

CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELEPHONE: 312/474-0448
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 9515 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Pseudomonas aeruginosa
STRAIN: Clinical Isolate P2-2
US-08-362-577C-13

Query Match 1.38; Score 45.4; DB 1; Length 9515;
Best Local Similarity 46.28; Pred. No. 0.033;
Matches 151; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

Qy 438 catcgtcagcggggaagatcgcccgaggagagctgcaactgcccgcgtccgagttcga 497
Db 802 CATGGCAAGCCGGTGTGGAGCGCTCGAATCATGATACCGCGCGCCACGCTTT 861

Qy 498 cgacacgttttccaagaattcgaggtgctcttctgcggccgctgacggtggcgacaa 557
Db 862 CGCCTGGTATGCGGAAGCCTCGAAGCTCTACGACAGGTGCGCGCGCCGACGA 921

Qy 558 gaaggtccgcggccctgatgacagagtgatcgagagttcaatacgtcagcggcag 617
Db 922 GACCCTGGCCACCATTACCCGCTGCCGTGGGGTGATCGCGCGGTGGTGGTGA 981

Qy 618 ccggggttcagagagcccgcccaaccccgcccgcccgcccgagagcagga 677
Db 982 CTTCGGCTGACATGCGCGCTGGAAGCTGCGCCCGCTGGCGCGGCACTCGT 1041

Qy 678 gctgtgcgagggccatgcgcaagtccttcttccagcccgccctgcgtcggtgctt 737
Db 1042 GGTGCTCAAGCGCGCGAGCAGTCGCCGTTCTCCGCCCTGGCGCTGGCGCT 1101

Qy 738 taggaagagctgcagatgggggct 764
Db 1102 GGAGCGGGGGTGGCGGAAGCGTGCT 1128

RESULT 7
US-08-920-828-13
; Sequence 13, Application US/08920828
; Patent No. 585398
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/920,828
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,577
FILING DATE: 27-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELEPHONE: 312/474-0448
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 9515 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Pseudomonas aeruginosa
STRAIN: Clinical Isolate P2-2
US-08-920-828-13

Query Match 1.38; Score 45.4; DB 2; Length 9515;
Best Local Similarity 46.28; Pred. No. 0.033;
Matches 151; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

Qy 438 catcgtcagcggggaagatcgcccgaggagagctgcaactgcccgcgtccgagttcga 497
Db 802 CATGGCAAGCCGGTGTGGAGCGCTCGAATCATGATACCGCGCGCCACGCTTT 861

Qy 498 cgacacgttttccaagaattcgaggtgctcttctgcggccgctgacggtggcgacaa 557
Db 862 CGCCTGGTATGCGGAAGCCTCGAAGCTCTACGACAGGTGCGCGCGCCGACGA 921

Qy 558 gaaggtccgcggccctgatgacagagtgatcgagagttcaatacgtcagcggcag 617
Db 922 GACCCTGGCCACCATTACCCGCTGCCGTGGGGTGATCGCGCGGTGGTGGTGA 981

Qy 618 ccggggttcagagagcccgcccaaccccgcccgcccgcccgagagcagga 677
Db 982 CTTCGGCTGACATGCGCGCTGGAAGCTGCGCCCGCTGGCGCGGCACTCGT 1041

Qy 678 gctgtgcgagggccatgcgcaagtccttcttccagcccgccctgcgtcggtgctt 737
Db 1042 GGTGCTCAAGCGCGCGAGCAGTCGCCGTTCTCCGCCCTGGCGCTGGCGCT 1101

Qy 738 taggaagagctgcagatgggggct 764
Db 1102 GGAGCGGGGGTGGCGGAAGCGTGCT 1128

RESULT 8
US-08-253-155A-9
; Sequence 9, Application US/08253155A
; Patent No. 5691147
; GENERAL INFORMATION:
; APPLICANT: Gyuris, Jeno
; APPLICANT: Draetta, Giulio
; TITLE OF INVENTION: CDK4 Binding Proteins
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA

```
;
;
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/253.155A
; FILING DATE: 02-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: MII-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-253-155A-9

Query Match 1.3% Score 45; DB 1; Length 8201;
Best Local Similarity 53.0%; Pred. No. 0.039;
Matches 96; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

Qy 2549 ttgacctggcggaactttctctacacaccatactctctgcccagcttgagcaggac 2608
Db 2133 TGGAGCTGAGGAGCACTCTCCGGAACCATGCTTTCTTAGGAYCGATGAGGCAAGC 2192

Qy 2609 agctatcgctttacacattttgaaggcctactctctagaccaggaagtgggatatt 2668
Db 2193 AGAGGGAACACTATTCATACATCTCCCTGGCTATTCCGGAGTATAACCCGGAGGTGGGCTACT 2252

Qy 2669 gcaaggtctagcttttagcaggacatttgctctctcatatagtaggaaggagcgt 2728
Db 2253 GCAGGGACCTTGAGCACCATCACCGCTTGTCTCTCTTTATCTGCTTGAGGAGGACGAT 2312

Qy 2729 t 2729
Db 2313 T 2313

RESULT 9
US-08-295-060-3
; Sequence 3, Application US/08295060
; Patent No. 5659123
; GENERAL INFORMATION:
; APPLICANT: VAN RIE, Jeroen
; APPLICANT: JANSSENS, Stefan
; APPLICANT: PERFEROEN, Marrix
; TITLE OF INVENTION: NEW DIABROTICA TOXINS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/295.060
; FILING DATE: 26-AUG-1994
```

```
;
;
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-052
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1957 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 13..1947
; US-08-295-060-3

Query Match 1.2% Score 41; DB 1; Length 1957;
Best Local Similarity 47.9%; Pred. No. 0.22;
Matches 113; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

Qy 488 ccgagttcgagacacgcttttccaagaagttcgaggtgctcttctgcggcggtgacgg 547
Db 1322 CCCAGACCTAGCACAGCANNCGCAACGCTGGCGCTGTGAGCTGGGACAGCATCGACCAGC 1381

Qy 548 tggcgacaaagaagctccgcgcctgatcgacgagtgcatcgagagttcaatcacg 607
Db 1382 TGCCTCCCGAGACACACCGAGGCCCTCTGGAGAGGGGTACAGCCACCAGCTGACTACG 1441

Qy 608 tcagcggcagcgggggtccgagagagcccccgcggcccccgcggcccatgccgcgcacag 667
Db 1442 TGATGTGCTTCTGTATGTCAGGGCAGCAGGGGACCATCCAGTGTGACTGGACCCACA 1501

Qy 668 ggagccaggagcctgtgcgcaggcccatgcgcaagtcttcttccccagccgcctg 723
Db 1502 AGAGCGTGGAGTCTTCTTCAACATGATCGACAGCAAGATCACCAGCTGCCCCCTG 1557

RESULT 10
US-08-858-003-2
; Sequence 2, Application US/08858003
; Patent No. 6060234
; GENERAL INFORMATION:
; APPLICANT: Katz, Leonard
; APPLICANT: Stassi, Diane L.
; APPLICANT: Summers Jr., Richard G.
; APPLICANT: Ruan, Xiaolan
; APPLICANT: Pereda-Lopez, Ana
; APPLICANT: Kakavas, Stephan J.
; TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Rd.
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858.003
; FILING DATE: 16-MAY-1979
```

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dianne Casuto
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 4952.US.P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847)-938-3137
; TELEFAX: (847)-938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1030 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-858-003-2

```

```

Query Match      1.1%; Score 40.8; DB 3; Length 1030;
Best Local Similarity 50.0%; Pred. No. 0.18;
Matches 102; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 440 tcctcagcgggggaagatcgcccgagagagagcagtcgactgcccgtccgagttcgacg 499
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 203 TCACACAGCGGCTCTGTTCCGCGGTGGAGACCTCCCTGTTCCGGTGTTCGAGGCCACG 262
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 500 acacgtttccaaagatcgagtgcttctgcgcccgcgtgacggtgcgcgacacaaga 559
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 263 GCTCGTCCCGACACTCATCGGCCACTCCATCGGCGAAGTGACCGGGCCACCTGG 322
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 560 aggtccgcggccctgacgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcg 619
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 323 CCGGGTCTCTCATCTGCGGGACGCTGCTGCTGCTGCCACCGCGCGCCTGATGC 382
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 620 gggggtcggagagcccccaccca 643
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 383 AGTCGCGCGCGCGCGCGCGCGA 406
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 11
US-09-762-166-2
; Sequence 2, Application US/09078166
; Patent No. 6083561
; GENERAL INFORMATION:
; APPLICANT: Katz, Leonard
; APPLICANT: Stassi, Diane L.
; APPLICANT: Summers Jr., Richard G.
; APPLICANT: Ruan, Xiaohan
; APPLICANT: Pereda-Lopez, Ana
; APPLICANT: Kakavas, Stephan J.
; TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Rd.
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/078,166
; FILING DATE: 16-MAY-1979
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dianne Casuto
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 4952.US.P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847)-938-3137
; TELEFAX: (847)-938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1030 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-078-166-2

```

```

Query Match      1.1%; Score 40.8; DB 3; Length 1030;
Best Local Similarity 50.0%; Pred. No. 0.18;
Matches 102; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 440 tcctcagcgggggaagatcgcccgagagagagcagtcgactgcccgtccgagttcgacg 499
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 203 TCACACAGCGGCTCTGTTCCGCGGTGGAGACCTCCCTGTTCCGGTGTTCGAGGCCACG 262
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 500 acacgtttccaaagatcgagtgcttctgcgcccgcgtgacggtgcgcgacacaaga 559
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 263 GCTCGTCCCGACACTCATCGGCCACTCCATCGGCGAAGTGACCGGGCCACCTGG 322
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 560 aggtccgcggccctgacgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcg 619
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 323 CCGGGTCTCTCATCTGCGGGACGCTGCTGCTGCTGCCACCGCGCGCCTGATGC 382
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 620 gggggtcggagagcccccaccca 643
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 383 AGTCGCGCGCGCGCGCGCGCGA 406
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 12
US-08-997-467-2
; Sequence 2, Application US/08997467
; Patent No. 6200813
; GENERAL INFORMATION:
; APPLICANT: Katz, Leonard
; APPLICANT: Stassi, Diane L.
; APPLICANT: Summers Jr., Richard G.
; APPLICANT: Ruan, Xiaohan
; APPLICANT: Pereda-Lopez, Ana
; APPLICANT: Kakavas, Stephan J.
; TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Rd.
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,467
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/858,003
; FILING DATE: 16-MAY-1997

```

ATTORNEY/AGENT INFORMATION:
NAME: Dianne Casuto
REGISTRATION NUMBER: P-40,943
REFERENCE/DOCKET NUMBER: 4952.US.P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (847)-938-3137
TELEFAX: (847)-938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1030 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-997-467-2

Query Match	1.1%	Score 40.8;	DB 4;	Length 1030;
Best Local Similarity	50.0%;	Pred. No. 0.18;		
Matches 102; Conservative	0;	Mismatches 102;	Indels 0;	Gaps
QY	440	tcgcgcagcggggaagatcgcccgagagagagctgcacatgcccgctccgagttcgcagc	499	
Db	203	TCACACAGGGGGCTCTGTTCGGCTGGAGACCTCCTCTGTTCGGCTGTTCGAGGCCACG	262	
QY	500	acacgtttttccaaagaattcgaagtgctcttctggcgccgctgaagtgggcgcaacaaga	559	
Db	263	GCCTCTGCCCGGACTACCTCATCGGCACATCCATCGGCAGTGAACGGCGGCCACCTGG	322	
QY	560	aggctccgcggccctgatcgcagcagtgcatcagagaagttcaatcagtcacgcggcagcc	619	
Db	323	CCGGGGTCTCTGATCTGTGGCGAGCGCTGGTCTCTGTCTGCCACCGCCGCTGATGC	382	
QY	620	gggggtccgagagccccccgccc	643	
Db	383	AGTCGGCCCGGGCCGCGCGCGCA	406	

RESULT 13
US-08-814-052-19
: Sequence 19, Application US/08814052
: Patent No. 6015783
: GENERAL INFORMATION:
: APPLICANT: von der Osten, Claus
: APPLICANT: Cherry, Joel R.
: APPLICANT: Bjornvad, Mads E.
: APPLICANT: Vind, Jesper
: APPLICANT: Rasmussen, Michael Dolberg
: TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING
: TITLE OF INVENTION: OR STAINS FROM CELLULOSIC FABRIC
: NUMBER OF SEQUENCES: 55
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 6015783o No. 6015783disk of No. 6015783th America, Inc.
: STREET: 405 Lexington Avenue, Suite 6400
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: Fastseq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/814,052
: FILING DATE: 06-MAR-1997
: CLASSIFICATION: 510
: ATTORNEY/AGENT INFORMATION:
: NAME: Lambiris, Elias J
: REGISTRATION NUMBER: 33,728
: REFERENCE/DOCKET NUMBER: 4684.204-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2249 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-814-052-19

	Query Match	1.1%	Score 40.6;	DB 3;	Length 2249;
	Best Local Similarity	48.9%;	Pred. No. 0.31;		
	Matches 109;	Conservative	0;	Mismatches 114;	Indels 0;
	Gaps				
Qy	474	gctgcactgcccgtccgagttcgacgacagcttttcccagaagttcggaggtgctctcttg	533		
Db	1234	GCTCTGGGGGGCTCCAGGAATCCCTACCGGCCGCATCTCCACTACGGCGGGGCCCC	1293		
Qy	534	cggccgctgtacggtgtggcgacaaagaaggtccgcccggccctgatcgacgagtgcatcga	593		
Db	1294	CGCGCGCCCGCCACGAGGAGGGCAAGGCCCGGTGCGACCACTGCTCTGGACCTCCC	1353		
Qy	594	gaagttcaatcacgtcacggcagcgggggtccgagagcccccgcccccaaccgcgccca	653		
Db	1354	CACCTTCAGCCCGTCTGTGGCCGGACGTGCCCTTGAGCGGCTTCGCCAAGCGGCCGA	1413		
Qy	654	tgcgcgccccacagggagccagagcctgtgcgagggcccatg	696		
Db	1414	CACACGCTCGAGCTCACCTCGACACACAGGGCAGCGCCCTG	1456		

```

RESULT 14
US-08-814-052-17
; Sequence 17, Application US/08814052
; Patent No. 6015783
; GENERAL INFORMATION:
; APPLICANT: von der Osten, Claus
; APPLICANT: Cherry, Joel R.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Vind, Jesper
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING
; TITLE OF INVENTION: OR STAINS FROM CELLULOSIC FABRIC
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6015783o No. 6015783disk of No. 6015783th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: PastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/814,052
; FILING DATE: 06-MAR-1997
; CLASSIFICATION: 510
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4684.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2279 base pairs
; TYPE: nucleic acid

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September .9, 2002, 16:41:06 : Search time 5191.84 Seconds
(without alignments)
9241.748 Million cell updates/sec

Title: US-09-762-311-4_COPY_176_3730

Perfect score: 3555

Sequence: 1 atggaacaaataacattcac.....tccacactgtccagccctt 3555

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estin.*
- 4: em_estnu.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_hic.*
- 9: gb_est1.*
- 10: gb_est2.*
- 11: gb_hic.*
- 12: gb_gss.*
- 13: em_gss_hum.*
- 14: em_gss_inv.*
- 15: em_gss_pln.*
- 16: em_gss_vrt.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
c					
1	833.6	23.4	1001	10	278359 HS278359 Hu
2	826.6	23.3	1051	10	BM479578 AGENCOURT
3	819.2	23.0	1007	10	BM016445
4	782.4	22.0	784	10	BM016445 603641252
5	741.8	20.9	835	10	BM016445 603641252
6	738.8	20.8	800	10	BM016445 603641252
7	737.4	20.7	1121	10	BM016445 603641252
8	729.8	20.5	1109	10	BM016445 603641252
9	722.2	20.3	952	10	BM016445 603641252
10	713.8	20.1	838	10	BM016445 603641252
11	705	19.8	807	10	BM016445 603641252
12	701	19.7	832	10	BM016445 603641252
13	689.8	19.4	851	9	AL570425
14	683.4	19.2	754	10	BE300005 600944594
15	682.6	19.2	892	10	BE561889 601346098
16	678.6	19.1	686	10	BE769758 602744658
17	675	19.0	1108	10	BG326728 602425604

18	671.8	18.9	978	10	BG029927
19	669.6	18.8	670	9	AL557623
c	669.4	18.8	857	9	AL544177
20	668.4	18.8	750	10	BI084159
21	668.4	18.7	795	10	BE560184
22	664.6	18.6	661	9	AL046787
23	661	18.6	661	9	AL046787
24	661	18.6	948	10	BE258116
25	659.2	18.5	672	10	BE675153
26	652.8	18.4	806	9	AM043925
27	648.8	18.3	665	9	AM173375
28	647.8	18.2	812	10	BI758489
29	641	18.0	1161	10	BM466841
c	637.8	17.9	773	10	BE561312
30	636.4	17.9	862	10	BI910369
31	636.4	17.9	660	10	BE276812
32	635.4	17.9	660	10	BE276812
33	632.4	17.8	761	10	BF797592
34	632	17.8	753	10	BF305442
35	630	17.7	722	10	BE560392
36	628	17.7	748	10	BG424359
37	624.4	17.6	626	9	AW772402
38	621.2	17.5	626	9	AW239183
39	612	17.2	623	10	BE884440
40	604.4	17.0	649	10	BE513375
41	597	16.8	605	9	AW732976
42	596.2	16.8	622	10	BE514108
43	592.4	16.7	774	10	BI763670
44	590.2	16.6	952	10	BG424757
45	589.4	16.6	758	10	BG166636

ALIGNMENTS

RESULT	1
278359/c	
LOCUS	1001 bp mRNA linear EST 28-JUL-1999
DEFINITION	HS278359 Human fetal brain S. Meier-Ewert Homo sapiens cDNA clone 3.114 (CEPH) 5', mRNA sequence.
ACCESSION	278359
VERSION	278359.1 GI:1495132
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	1 (bases 1 to 1001)
AUTHORS	Neri, C., Albanese, V., Lebre, A. S., Holbert, S., Saada, C., Bouqueleret, L., Meier-Ewert, S., Legall, I., Millasseau, P., Bul, H., Giudicelli, C., Massart, C., Guillou, S., Gervy, P., Poullier, E., Rigault, P., Weissenbach, J., Lennon, G., Chumakov, I., Dausset, J., Lehrach, H., Cohen, D. and Cann, H. M.
TITLE	Survey of CAG/CTG repeats in human cDNAs representing new genes: candidates for inherited neurological disorders
JOURNAL	Hum. Mol. Genet. 5 (7), 1001-1009 (1996)
MEDLINE	96414310
COMMENT	Contact: Neri C. Fondation Jean Dausset - CEPH 27 Rue Juliette Dodu, 75010 Paris, France Related sequence: 278360 5'-sequence (upper strand). Location/Qualifiers 1. .1001 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="3.114 (CEPH)" /clone_lib="Human fetal brain S. Meier-Ewert" /tissue_type="brain" /dev_stage="fetus" /note="cDNA library of S. Meier-Ewert, Max Planck Inst.f. Mol. Genetics, Berlin, FRG"
BASE COUNT	237 a 229 c 231 g 272 t 32 others
ORIGIN	


```
Qy 2811 gtaccagctctcaggttgcttcattgattaccacagagacctctacaatcacctggagga 2870
|||||
Db 361 GTACCAGCTCTCGAGGTGCTTCATGATTACCACAGAGACCTCTACAATCACCTGGAGGA 420
Qy 2871 gcacgaatcgcccccagctctacgtgcccctggcttccctcacaatgcttgcctcaea 2930
|||||
Db 421 GCACGAGATCGGCCCCAGCCTCTACGCTGCCCTGGTTCCCTCACCATGTTTGCCTCACA 480
Qy 2931 gtcccgctgggtatctgacagagctttgatgatattttcttcagggaaacaggt 2990
|||||
Db 481 GTTCCCGCTGGGATTCGTACCGAGAGCTTTGATATGATTTTCTTCAGGGAACAGAGGT 540
Qy 2991 catatttaagtggctttaagtctgttgggaagccataaagcccttgattctgcagcatga 3050
|||||
Db 541 CATATTTAAAGTGGCTTTAAAGTCTGTGGGAAGCCATAAAGCCCTTGATTCTGCAGCATGA 600
Qy 3051 aaacctagaaccatagttgacittataaaagcagcctacccaaacctggcttggtata 3110
|||||
Db 601 AAACCTAGAACAACATAGTTGACTTATATAAAAGCAGCTTACCCAAACCTTGGCTTGATA 660
Qy 3111 gatggaagaccatcaatcaggtattgaaatggacatcgctaaacagttacaagctta 3170
|||||
Db 661 GATGGAAGAACCATCATCAGTATTTGAATGGACATCGCTTAACAGATTACAGCTTA 720
Qy 3171 tgaagttgattaccagctctcctcaagaagaacttatcgattcctctcctctcagtgacaa 3230
|||||
Db 721 TGAAGTTGACTACCAGCTCTCTCAAGAAGAACTTATCGATTCTCTCTCTCAGTGACAA 780
Qy 3231 ccaagaatggataaattagagaaaacacagcagcgtcttacgaaacagacaccttgacct 3290
|||||
Db 781 CC-AAGAATGGATTAATAGAGAAACCAACAGCAGGCTTACGCAACAGAAACCTTGACCC 839
Qy 3291 ccttgaaagcttcaggt-ggcaaatggttaggagatcaaaagccttg--aggccaccattga 3347
|||||
Db 840 TCCTTGACATCTCAGGTGGGCCAATGGTAAGATCAAAAGCCCTTGAGGCCCTCCCATGGA 899
Qy 3348 gaagct-cctgagcagtgagagcaagctgaa 3377
|||||
Db 900 AAAGCTCCCTGAACAGGAGACCAAGCCATA 930

RESULT 3
BM016445
LOCUS
DEFINITION
603641252F1 NIH_MGC_87 Homo sapiens CDNA clone IMAGE:5417148 5',
mRNA sequence.
ACCESSION
BM016445
VERSION
BM016445.1 GI:16530799
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1007)
NIH-MGC http://mgi.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
JOURNAL
Contact: Robert Strausberg, Ph.D.
COMMENT
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12064 row: h column: 13
High quality sequence stop: 822.
Location/Qualifiers
1..1007
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5417148"
```

```
BASE COUNT 274 a 246 c 255 g 231 t 1 others
ORIGIN

/clone_lib="NIH_MGC_87"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: breast; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dr primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."

Query Match 23.0%; Score 819.2; DB 10; Length 1007;
Best Local Similarity 95.1%; Pred. NO. 8.9e-194;
Matches 910; Conservative 0; Mismatches 39; Indels 8; Gaps 6;

Qy 2532 gcagcatgaggtctctattgaccttgggggaacctttctacacacccatactctctgc 2591
|||||
Db 1 GCAGCATGCGATTCTTATTGACCTTGGCGAAGCTTTCTCTACACACCATACTTCTCTGC 60
Qy 2592 ccagcttgagcagcagcagctatcgctttacaacatttgaagcctcactcactctaga 2651
|||||
Db 61 CCAGCTTGGAGCAGGACAGCTATCGCTTTTACAACATTTTGAAGCCCTTACTCATCTTCTAGA 120
Qy 2652 ccaggaagtgaggatattgccaaggtctcagctttgttagcaggcatatttgccttctcatat 2711
|||||
Db 121 CCAGGAAGTGGGATATTGCCAAGTCTCAGCTTTGTAGCAGGCANTTTGCTTCTTCATAT 180
Qy 2712 gagtgggaagagggcgttataaagtctcaagttctcagttgttgacatggggctgggaa 2771
|||||
Db 181 GAGTGAGGAAGAGGCGCTTAAATATGCTCAAGTTTCTGATGTTTGACATGGGGTGGCGAA 240
Qy 2772 acagttatcgccagcagcagctattttacagatccagatgacagctctcagagtgct 2831
|||||
Db 241 ACAGTATCGGCCAGACATGATTATTTTACAGATCCAGATGTACAGCTCTCGAGGTTGCT 300
Qy 2832 tcattgattaccacagagacctctacaatcacctggaggagcagcagatcgccccagcct 2891
|||||
Db 301 TCATGATTACCACAGAGACCTCTACAATCACCTGGAGGACACGAGATCGGCCAGCCT 360
Qy 2892 ctacgtgccccctgggttctcctaccatggttgctcagagttccccgctgggattcgtagc 2951
|||||
Db 361 CTACGTGCCCCCTGGTCTTCTCACCATGTTTGCTTCCAGTTCCTCCGCTGGGATTCGTAGC 420
Qy 2952 caagctcttgatgatattttcttcagggaacagaggtcatatttaaagtggctttaa 3011
|||||
Db 421 CAGAGCTTTTGATATGATTTTCTTCAGGGAACAGAGGTCATATTTAAAGTGGCTTTAAG 480
Qy 3012 tctgttggaagccataagcccttgattctgcagcatgaaacctagaaacctagatttga 3071
|||||
Db 481 TCTGTGGGAAGCCATAAGCCCTTGATTCTGCAGCATGAAACCTAGAAACCATAGTTGA 540
Qy 3072 ctttataaaagcagcagcctaccacccctggcttggttacagatggaaaagaccataatca 3131
|||||
Db 541 CTTTATAAAAGCAGCTACCCAACTTGGCTTGGTACAGATGGAAAGACCACATCAATCA 600
Qy 3132 ggtattgaaatggacatcgctaaacagttacagcttatgaagttgagtcacacgtcct 3191
|||||
Db 601 GGTATCTGAATGGACATCGCTAAACAGTTTACAGCTTATGAAGTTGAGTACCACGCTCT 660
Qy 3192 tcaagaagaacttatcgattcctcctctcagtcagcaacccaaagaatgataa-attag 3250
|||||
Db 661 TCAAGAAGAAGCTTATCGATTCTCTCTCTCAGTACAAACCAAGAAATGATTAACCTTAG 720
Qy 3251 agaaaacccacagcagcagcttacgcaaacacagaccccttgacctcttgacagttgca-agg 3309
|||||
Db 721 AGAAAACCAACAGCAG-TTACGCAAAACAGAACCTTGGACTCTTGAACAGTTGCAGGGTG 779
Qy 3310 gcaaatggttagatccaaagccttgaggccaccattgagaa--gctcctgagcagtgaga 3367
|||||
Db 780 GCAAAATGTTAGATCCAAAGCCTTGAGGCCCTTGAAGGAGAAAGCTCCCGGAGCAGTGAGA 839
Qy 3368 gcaag--ctgaagcaggccatgcttacc-ttagaactggagcggtcgccctgctgcaga 3424
```

```

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 840 GCAAGCTGGAGCAGGCATGCTTACCTTTAGAACTGGAGCGCTCGCCCTGCTGCAGA 899
QY 3425 cgtgagagactcgcgcgagcgagcgagagccagagccagagcgagcgagcctgaagtca 3481
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 900 CGGTGGAGAACTCGCGCGGGGAACCGAAGCCAGCCAGCGAGCGGCGGAGTGCA 956

RESULT 4
BG828092 784 bp mRNA linear EST 22-MAY-2001
LOCUS 60275346F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4906191 5',
DEFINITION mRNA sequence.
ACCESSION BG828092
VERSION BG828092.1 GI:14175679
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 784)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI808 row: n column: 16
High quality sequence stop: 784.
Location/Qualifiers
1..784
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4906191"
/clone_lib="NIH_MGC_17"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT 236 a 190 c 182 g 176 t
ORIGIN

Query Match 22.0%; Score 782.4; DB 10; Length 784;
Best Local Similarity 99.9%; Pred. No. 1.3e-184;
Matches 783; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1859 ggaactatgaggtatcactcagcagacagacacccctcagacgaagaagcatttg 1918
Db 1 GGAAACTTATGAGGTATCACTCAGTGAGCAGACGCGCTCATGAACGAAGGACTTTG 60

QY 1919 aaacaaagcaaacatcttggtgattcttggtggactcctgtgaagcccgaggcatt 1978
Db 61 ATCCAAAGCAACCATCTTGTGATTCTGGTGGACTCCCTGTGAAGACCGGAGGCAATT 120

QY 1979 cctgaggcagcagatattctcccgagtagccaccccgaggaagcgctgcattctcca 2038
Db 121 CTGGAGGCGCAGCATATTCTCCGAGTAGCCACCCCGCAGGAAGCGTGCAGATTCTCCA 180

QY 2039 gcagatatgaagattattcagagctgggagagctccccccagatcctcttagaaccag 2098
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

```

Db 181 GCAGATATGAAGATTATTTCAGAGCTGGGAGAGCTTCCCCACGATCTCCTTTAGAACGAG 240
QY 2099 ttgtgaagatgggccccttttgccccccacagagagaaagaaagagacatctcgtgagc 2158
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 241 TTTGTGAAGATGGGCCCTTTGGCCCCCCCACAGAGAGAAAGAAAGAGACATCTCTGTGAGC 300
QY 2159 tccgagagctgtggcaaaaggctattcttcaacagatactgctgcttagaagtggaagg 2218
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 301 TCCGAGAGCTGTGGCAAAAGGCTATTCTTCAACAGATACTGCTGTTAGAAATGGAGAAAG 360
QY 2219 aaaaatcaagaactccaaagcctctgaaaatgatttgcatacaacagcgctgaagcctcgatt 2278
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 361 AAAATCAGAAGCTCCAGCCCTCTGAAAATGATTCTGTAACAAGCGCTGAAGCTCGATT 420
QY 2279 atgaagaaattactccctgtctttaaagaagtaactacagtggtggaaaagatgcttagca 2338
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 421 ATGAAGAAATTACTCCCTGTCTTAAAGAAAGTAACATACAGTGTGGGAAAGATGCTTAGCA 480
QY 2339 ctccagggaagatcaaaaataagtttgacatggaaaataatgcactcgcgtgttgaggcaag 2398
Db 481 CTCCAGGAAGATCAAAAATTAAGTTTGATCTGGAATAATGCACCTCGGCTGTGGGCAAG 540
QY 2399 gtgtgccacgtcatcacccagagtgaaaatcttgaaaattcttagctgagcaattccacctta 2458
Db 541 GTGTGCCACGTCATCCAGCGAGTGAAATCTGGAAATTTCTAGCTGAGCAATTCACCTTA 600
QY 2459 aacaccagtttcccagcaaacagcagcgaaggatgtgccatacaaaaagactcttaagc 2518
Db 601 AACACCAGTTTCCAGCAACAGCAGCAGCAAGGATGTGCCATACAAAGAACTCTTAAAGC 660
QY 2519 agctgacttcccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2578
Db 661 AGCTGACTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 720
QY 2579 catacttctgcccagcgttgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2638
Db 721 CATACTTCTCTGCCCGAGCTTGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 780
QY 2639 actc 2642
|||||
Db 781 ACTC 784

RESULT 5
BG759286 835 bp mRNA linear EST 15-MAY-2001
LOCUS 602710975F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4851090 5',
DEFINITION mRNA sequence.
ACCESSION BG759286
VERSION BG759286.1 GI:14069939
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 835)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI693 row: f column: 19
High quality sequence stop: 766.
Location/Qualifiers
1..835
/organism="Homo sapiens"

```

```
/db_xref="taxon:9606"  
/clone_lib="IMAGE:4851090"  
/tissue_type="Primary B-cells from tonsils (cell line)"  
/lab_host="DH10B (phage-resistant)"  
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;  
Site_2: EcoRI; CDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH-MGC Library." 171 t  
BASE COUNT 220 a 230 c 214 g 171 t  
ORIGIN
```

```
Query Match 20.98; Score 741.8; DB 10; Length 835;  
Best Local Similarity 98.8%; Pred. No. 1.9e-174;  
Matches 768; Conservative 0; Mismatches 7; Indels 2; Gaps 2;  
  
QY 1591 gatagtccctgtctagatcaatgaacacacagcaagagccatctgtgtgaaag 1650  
Db 1 GATAGCTCCCTGCTAGTACATTAAGTAACACACGAAAGAGCCATCTGTGTGAAAG 60  
QY 1651 gaggccttgcctatctctgagagctcttaagctctcgtcctcgtgaggacctgtcc 1710  
Db 61 GAGGCCCTTGCCCATCTCTGAGAGCTCCTTAAGCTCCTCGCTCCTCGGAGGACCTGTC 120  
QY 1711 agtgactcagagatctctccagaagagccagctcgtgtgcgccagcaggcttc 1770  
Db 121 AGTGACTCGGAGAGTATCTCCCAAGAGAGCCAGCTCCGTGTGCGCCCGAGGCGCTTC 180  
QY 1771 aggaggcagcaaacacccctgagtcacttcccatcgaatgccaggaaacctccacaacct 1830  
Db 181 AGGAGGCGAGCAACACCCCTGAGTCACTTCCCATCGAATGCCAGGAACCTCCACACCT 240  
QY 1931 gcccggggtcccgggggttgcgaagaaacttatgaggtatcactcagtgagca 1890  
Db 241 GCCCGGGGTCCCGGGGGTTCGCAAGGAAACTTTATGAGGTATCACTCAGTGAGACA 300  
QY 1891 gagcgcctcatgaacaaaggacttgaaatccaaagcaaacatctgtgtattcgt 1950  
Db 301 GAGACGCTCATGAACAAAGGACTTTGAATCCAAAGCAACCATCTGTGTGATCTGGT 360  
QY 1951 gggactcctgtgaagaccgagagcattcctggaggcagcagatattcctccagtagcc 2010  
Db 361 GGGACTCTGTGAAGACCCGGAGCATCTCTGGAGGCGAGAGATATTCCTCGAGTAGCC 420  
QY 2011 acccgcagagcgtcgtgattcttccagcagatataagattattcagagctgggagag 2070  
Db 421 ACCCGCAGAGGCGTGCAGATTCTTCAGCAGATATGAAGATTATTCAGAGCTGGGAGAG 480  
QY 2071 ctccccccagatctctttagaccagttgtgaagatggcccttggccccccacca 2130  
Db 481 CTTCCCCCAGATCTCTTTAGAACACAGTTTGTGAAGATGGGCCCTTTGGCCCCCACCAC 540  
QY 2131 gaggaaagaaaaggacatctcgtgagctcgcagagcgtgtggcaaaaggctattcttcaa 2190  
Db 541 GAGGAAAGAAAAGGACATCTCGTGAGCTCCGAGAGCTGTGGCAAAAGGCTATTCTTCAA 600  
QY 2191 cagatactgctgttagaattgggaagaaaatacagaagctccaagcctctgaaatgat 2250  
Db 601 CAGATACTGTGTGTAGAAATGGAGAGAAATCAGAAAGCTCCAAGCTCTGAACATGAT 660  
QY 2251 ttgctgaacagcgcctgagctcgattatgaagaaattactcctgtcttaagaagatga 2310  
Db 661 TTGCTGAACAGGCCCTGAGCTTCGATTATGACGAATTAATCTCCTGTCTTAAGAAGTA 720  
QY 2311 actacagtgt-gggaaaaagatgctta-gcactccaggaagatcaaaaaataagttg 2365  
Db 721 ACTACAGTGTGGGGAACAGATGCTTAGGCACCTCCAGGAGCATCAACAATCAGGTTTCG 777
```

RESULT 6

BG746377

LOCUS

DEFINITION

BG746377 BG746377 800 bp mRNA linear EST 15-MAY-2001

602703644F1 NIH_MGC_15 Homo sapiens CDNA clone IMAGE:4856891 5',

mRNA sequence.

ACCESSION

BG746377

VERSION

BG746377.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 800)

AUTHORS

NIH-MGC

TITLE

http://mgs.nci.nih.gov/

JOURNAL

National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-f@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: NIH Intramural Sequencing Center

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1C1708 row: h column: 12

High quality sequence stop: 762.

FEATURES

source

1..800

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="IMAGE:4856891"

/tissue_type="NIH_MGC_15"

/lab_host="DH10B (phage-resistant)"

/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:

EcoRI; CDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCACGAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 213 a 218 c 203 g 166 t

ORIGIN

Query Match 20.88; Score 738.8; DB 10; Length 800;

Best Local Similarity 98.8%; Pred. No. 1.1e-173;

Matches 787; Conservative 0; Mismatches 7; Indels 4; Gaps 4;

QY 1502 agagatcttttaacagagctcttagaagattttgtcccggggtaataaagccagagcc 1561

Db 2 AGAGATCTTTAACAGAGCTTTAGAAAGATTTTGTCCGGGGTAATAAGCCAGAGGCC 61

QY 1562 tgcaggaaactccatcagtggtgagctgagtagctccctgtctagtagtaagtaaca 1621

Db 62 TGCAGGAACACTCCATCAGTGTGATCTGGATAGTCCCTGTCTAGTACATTAGATAACA 121

QY 1622 ccagcaagagccatctgtgtgaaagagggccttgccatctctgagagctccctta 1681

Db 122 CCAGCAAGAGCCATCTGTGTGAAAGAGGCGCTTGCCCAATCTCTGAGAGCTCCTTTA 181

QY 1682 agctcctcggtcctcctcgaggagcctgtccagtgactcggagagtcattctccagaagagc 1741

Db 182 AGCTCTCGGCTCCTCGAGGACCTGTCCAGTGACTCGGAGAGTCATCTCCAGAGAGC 241

QY 1742 cagctcgtgtgcgccccagcagccttcagggagcagcaaacacctgagtcacttc 1801

Db 242 CAGCTCGGCTGTGGCCCCCAGCAGCCCTTCAGGAGCGGCAACACCCCTGAGTCACTTC 301

Db	433	AGCAGGCCCTTCAGAGGCGGAGCAAAACCCTGAGTCACTTCCCCATCGAATGCCAGAAC	492		
QY	1820	ctccaaacctgccgggggtcccccgggggttcgcgaaggaaccttatgaggtacct	1879		
Db	493	CTCCACAACCTGCCGGGGGTCCCGGGGGTTCGCAAGGAACTTATGAGGTATCACT	552		
QY	1880	cagtgcacagagacgcctcatgaacgaaggaacctttgaatccaaagcaaacattctg	1939		
Db	553	CAGTGCACACAGAGACGCCCTCATGACGAAGGACTTTGAATCCAAAGCAAGCATCTG	612		
QY	1940	gtatcttggtgagactcctgtgaagaccggagcgattcctgagggcagcagatatcc	1999		
Db	613	GTGATCTGTGTGGACTCCTGTGAAGACCCCGGAGCATCTCTGGAGGCACGATATCC	672		
QY	2000	tccgagtgcacacccgcgaagaagcgtgcgattcttccagcagatatgaagattattcag	2059		
Db	673	TCCGAGTAGCACCCCGCAGAAAGGCGGCGATCTTCCACGACAGATGAGATTATTCG	732		
QY	2060	agct-gggagagattccccacagatc-tcctttagaaccagtttgt-gaagatggggcct	2116		
Db	733	AGCTGGGAGAGGCTCCCGCCCAACTTCCCTTAGAACCACTTGTGTGAAGATGGGGCC	792		
QY	2117	ttggccccccacc	2129		
Db	793	TTTGGGCGCCTCC	805		
RESULT 9					
BE299948					
LOCUS	600944494F1	NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960782 5',	linear EST 20-JUL-2000		
DEFINITION	mRNA sequence.				
ACCESSION	BE299948				
VERSION	BE299948.1	GI:9183696			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi.				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 952)				
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/				
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov Plate: LCM54 row: k column: 23 High quality sequence stop: 709. Location/Qualifiers 1. .952 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2960782" /clone_lib="NIH_MGC_17" /tissue_type="rhodomyosarcoma" /lab_host="DH10B (phage-resistant)" /notes="Organ: muscle; Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."				
BASE COUNT	249 a	240 c	250 g 212 t		
ORIGIN	1 others				

```
Query Match      20.3%; Score 722.2; DB 10; Length 952;
Best Local Similarity 97.2%; Pred. No. 1.7e-169;
Matches 766; Conservative 0; Mismatches 19; Indels 3; Gaps 3;

QY 1879 tcagtgcagacagagcgcctcatgaacgaagaggaacttgaatccaaagcaaacattctt 1938
DB 1 TCAGTGAGCAGACAGAGCGCTCATGAACGAAGAGGACTTGAATCCAAAGCAACCATCTTT 60
QY 1939 ggtgattctggtgggactcctgtgaagaccgagagcattctctgagcgagcagatattc 1998
DB 61 GGTGATTCTGTGGGACTCCTGTGAAGACCCGGAGGCATCTCTGGAGCGAGCAGATATTC 120
QY 1999 tctccgagtagccaccgccgcagagggcgtgctgattcttccagcagatatgaagattattca 2058
DB 121 CTCGAGTAGTACCACCCCGCAGAAAGGGCTGCATCTCTCCAGCAGATATGAAGATTATTC 180
QY 2059 gagctgggagagcttccccccagatctctcttagaaccagtttgaaagatgggcccctt 2118
DB 181 GAGCTGGGAGAGCTTCCCCCAGCATCTCTTTAGAACCAAGTTGTGAAGATGGGCCCTTT 240
QY 2119 ggcgcgcaccagagaaagaaagagacatctcgtgagctccgagagctgtggcaaaag 2178
DB 241 GCGCCGCCACCCAGAGAAAGAAAGAGACATCTCTGAGCTCCGAGAGCTGTGGCAAAAG 300
QY 2179 gctattcttcaacagatactgctgcttagaattggaagaaagaaatcagaagctccaagcc 2238
DB 301 GCTATTCTTCAACAGATACGTGCTGTAGATGGAGAGGAAATCAGAAGCTCCCAAGCC 360
QY 2239 tctgaaattgatttgcgaagacgcctgaagctcgattatgaagaaattactcctgt 2298
DB 361 TCTGAAATATGATTGTCTGAACAAAGCGCTGAAGCTCGATTATGAAGAAATTAATCCCTGT 420
QY 2299 cttaagaagtaactacagctgtgggaaagatgcttagcactccaggaagatcaaaatt 2358
DB 421 CTTAAGAAGTAACTACAGTGTGGGAAAGATGCTTTAGCACTCCAGAGAGATCAAAATT 480
QY 2359 aggtttgacatggaaaaatgcactggctgtgtgggcaagggtgtgccagctcatcaca 2418
DB 481 AAGTTTGACATGGAAAAATGCACTCGGCTG-TGGGCAAGGTGTGCCAGCTATCACCGA 539
QY 2419 ggtgaacttgaaatttctagctgagcaattccaccttaaacaccagtttcccgacaa 2478
DB 540 GGTGAATCTGGAAATTTCTAGCTGAGCAATTCACCTTTAAACACCAGTTTCCACGAAA 599
QY 2479 cagcagcaaaagatgtgccatacaagaactcttaagcagctgacttcccagcagcat 2538
DB 600 CAGCAGCCAAAGGATGTGCCATACAAAGAACTCTTAAGCAGCTGACTTCCCAGCAGCAT 659
QY 2539 gcaattctta-ttgaccttggcggaaccttctctacacaccatactctctgtgccagct 2597
DB 660 GCGATTCTTATTGTGACCTGNGCGAACCTTTCTCTACACACCCTACTTCTCTGCCAG-T 718
QY 2598 tgagcagacagctatcgtcttacaacattttgaagcgctactcacttctagaccagga 2657
DB 719 TGAGCAGCAGACAGCTATCGCTTTACAAACATTTGTGAGGCGCTACTCCACTCTTAGACAGGA 778
QY 2658 agtgggat 2665
DB 779 GTGGGTAT 786

RESULT 10
BI907513
LOCUS
DEFINITION BI907513 838 bp mRNA linear EST 16-OCT-2001
603065640F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5214945 5',
mRNA sequence.
ACCESSION BI907513
VERSION BI907513.1 GI:16170347
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE 1 (bases 1 to 838)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9pbbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11539 row: o column: 10
High quality sequence start: 4
High quality sequence stop: 780.

FEATURES
source
1..838
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5214945"
/clone_lib="NIH_MGC_118"
/tissue_type="leukocyte"
/lab_host="DH10B"
/Note="Vector: PCMV-SPORT6; Site_1: NotI; Site_2: EcoRV
(destroyed); RNA source leukocytes from anonymous pool of
non-activated adult donors. Library is oligo-dT primed
and directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
1.2-3.3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 027. Note:
this is a NIH_MGC Library."
BASE COUNT 237 a 199 c 191 g 211 t
ORIGIN
```

```
Query Match      20.1%; Score 713.8; DB 10; Length 838;
Best Local Similarity 97.0%; Pred. No. 2e-167;
Matches 812; Conservative 0; Mismatches 17; Indels 8; Gaps 8;

QY 2226 gaagctcaagcctctgaaatgatttgcgaacagccctgaagctcgattatgaaga 2285
DB 9 GAAGCTCCAAGCCTCTGAAATGATTGTCTGAACAAAGCCCTGAAGCTCGATTATGAAGA 68
QY 2286 aattactcctctcttaaagaagtaactacagctgtgggaaaaagatgcttagcactccagg 2345
DB 69 AATTACTCCTCTCTTAAAGAAGTAACACTACGTGTGGGAAAAGATGCTTAGCACTCCAGG 128
QY 2346 aagatcaaaaaataagtttgacatggaaaaaataagcactggctgttgggcaagggtgccc 2405
DB 129 AAGATCAAAAATTAAGTTTGACATGGAAAAATGCACTCGGCTGTGGGCAAGGTGTGCC 188
QY 2406 acgtcatcacccagctgaaatctgaaatttctagctgagcaattccaccttaaacacca 2465
DB 189 AGCTCATCACCCAGGTGAAATCTGGAAATTTCTAGCTGAGCAATTCACACTTAAACACCA 248
QY 2466 gttccccagcaaacagcagccaaagatgtgccatacaagaactcttaaacagcgtgac 2525
DB 249 GTTCCCAGCAAAACAGCAGCCAAAGGATGTGCATACAAAGAACTCTTAAGCAGCTGAC 308
QY 2526 tccccagcagcagtgagattctattgaccttggggaaaccttctcacacccactt 2585
DB 309 TTCCCAGCAGCATGGCGATTCTTA-TGACCTTGGCGAACTTCTCCATACACACCATACT 367
QY 2586 ctctgcacagcttgagcagcagcagctatcgctttacaacattttgaagcgctactcact 2645
DB 368 CTCTGCCCAGCTTGGAGCAGGACGACTATCGCTTTTACAACA-TTTGAAGCGCTTACTCACT 426
QY 2646 tctagaccaggaagtgggatatgtgcaaggtctcagctttagcagggatttggcttct 2705
DB 427 TCTAGACCAGGAAGTGGGATA-TGCCAAGGTCTCAGC-TTGTAGCAGGCA-TTTGTCTTCT 483
```


Qy 2706 tcataatgaggaagcgggttttaaaatgctcaagtttctgtatgtttgacatgggct 2765
 Db 484 TCATATGAGTGAGGAGAGGCGTTAAATGCTCAAGTTCTGAG-TTGACATGGGGCT 542
 Qy 2766 gcggaaacagtcagcggcagacatgattttacagatccagatgtaccagctctcgag 2825
 Db 543 GCGAAACAGTATCGGCGAGACATGATATTTTACAGATCCAGATGTACCAGCTCTCGAG 602
 Qy 2826 gtgtctcatgattaccacagagaccctctacaatcaactgagagacacagatcgccc 2885
 Db 603 GTTGCTTCATGATTAACACAGAGACCTCTACAATCACCTGGAGGAGCAGAGATCGGCC 662
 Qy 2886 cagcctctacgtc-gccccctggctctcaccatgttgcctcaccagttcccgctggat 2944
 Db 663 CAGCCTCTACGCTGCCGCCCTGGTCTCACCATGTTTGGCTCACATTCGCCGTGGGAT 722
 Qy 2945 tcgtagccagagtcgttgatgatttttcttcaggggaacagaggtcatatttaagtgg 3004
 Db 723 TCGTAGCCAGAGCTTTTGATGATGATTTTCTTTCAGGGAACAGAGGTCATATTAAAGTGG 782
 Qy 3005 ctttaagtctgtgggaagccataagcccttgattctgcagcatgaaacccatagaaa 3061
 Db 783 -TTTAAATCTGTGGGAGGCGATAAGCCCTGATTCTGCAGCATTTGAAACCTAGAAA 838

RESULT 11
 BG120498 807 bp mRNA linear EST 30-JAN-2001
 LOCUS 602346815F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:441923 5',
 DEFINITION mRNA sequence.
 ACCESSION BG120498
 VERSION BG120498.1 GI:12614007
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 807)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
<http://image.llnl.gov>
 Plate: LLAM10213 row: n column: 04
 High quality sequence stop: 682.
 Location/Qualifiers
 1..807
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:441923"
 /clone_lib="NIH_MGC_90"
 /tissue_type="adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.7 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."
 BASE COUNT 231 a 197 c 189 g 190 t
 ORIGIN

Query Match 19.8%; Score 705; DB 10; Length 807;
 Best Local Similarity 96.1%; Pred. No. 3.le-165;
 Matches 766; Conservative 0; Mismatches 25; Indels 6; Gaps 4;

Qy 1992 gatattctccagtagccaccccgagaaaggcgtgcgattcttccagcagatatgaaga 2051
 Db 1 GATATTCTCCGAGTAGCCACCCCGAGAAAGCGTGGATCTTCCAGCAGATATGAGA 60
 Qy 2052 ttattcagagctgggagagcttccccccacgatctcctttagaaccagttgtgaagtgg 2111
 Db 61 TTATTACAGAGCTGGGAGAGCTTCCCCCACCAGATCTCCTTTAGAACCACTTTGTGAAGATGG 120
 Qy 2112 gccctttggccccccacagagagaaagaaagacatctcgtgagctcggagagagctgtg 2171
 Db 121 GCCCTTTGGCCCCCACCAGAGGAAAGAAAGAGACATCTCGTGAGTCCGAGAGCTGTG 180
 Qy 2172 gcaaaaggctattcttcaacagatcactgctgtcttagaatgagaagaaatcagaagct 2231
 Db 181 GCAAAAGGCTATTCTTCAACAGATACACTGCTTGTAAGTGGAGAGAAATACAGAGCT 240
 Qy 2232 ccaagcctctgaaatgatttgcctgaacaagcgctgaagctcgattatgaagaaataac 2291
 Db 241 CCAAGCCTCTGAAATGATTGCTGGAACAAGCGCTCAAGCTCGATTATGAAGAAATAC 300
 Qy 2292 tcctgtcttaagaagaactaactacagtggtgggaaagatccttagcactccaggaagatc 2351
 Db 301 TCCCTGTCTTAAAGAAAGTAACACTACAGTGTGGGAAAGATGCTTAGCACTCCAGGAAGATC 360
 Qy 2352 aaaaattaaatttgacatggaaaaatgcactcgctgttgggcaaggtgtgccacgtca 2411
 Db 361 AAAAAATTAAAGTTGACATGGAATAATGCACCTCGGCTGTTGGCAAGGTGTGCCACGTCA 420
 Qy 2412 tcaccgaggtgaaatctggaaattcttagctgagcaattccaccttaaacacacagttcc 2471
 Db 421 TCACCGAGGTGAAATCTGAAATTTCTAGCTGAGCAATTCACCTTAAACACCACTTTCC 480
 Qy 2472 cagcaaacacagcagcaaaagatgtgccatacaagaactttaagcagctgactccca 2531
 Db 481 CAGCAAAACAGCAGCAAAAGGATGCGCATCAAAAGAACTCTTAAAGCAGCTGACTTTCCCA 540
 Qy 2532 gcagctgcgattcttattgaccttggcgaaacctttccacaca-cccatacttctctg 2590
 Db 541 GCAGCATGCGATCTTATTGACCTTGGCGAAGCTTTCTTACACAGCCCATACTTCTCTG 600
 Qy 2591 ccagcttgagcagcagcagctatcgctttacaacatttggaagcgctactactcttag 2650
 Db 601 CCCAGCTTGGAGCAGGACAGCTATCGCTTTTACAACTTTTGAAGGCTTACTACTTTCTAG 660
 Qy 2651 accaggaagtggtatattgcc--aaggtctcagcttt-gtagcagcattttgtctcttc 2707
 Db 661 ACCAGGAAGTGGGATATGCCCCAAAGTCTTCAGCTTTGGTGGCAGGCTCTTGTCTTCTC 720
 Qy 2708 atatgagtggaggaagcggctttaaagtctcaagtttctgattgttgacatggggctgc 2767
 Db 721 ATATGAGTGACGAGAGGCGTTAAATGCTCCA--GTTCCGATGTTTGACATGGGCCGCG 778
 Qy 2768 ggaacagtatcggcca 2784
 Db 779 GGAACCAAGTTCCGGCAA 795

RESULT 12
 BM014940 832 bp mRNA linear EST 30-OCT-2001
 LOCUS 603640877F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5416818 5',
 DEFINITION mRNA sequence.
 ACCESSION BM014940
 VERSION BM014940.1 GI:16529294
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 832)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: DCTD/DNP
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 found through the I.M.A.G.E. Consortium information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM12063 row: j column: 19
 High quality sequence stop: 811.
 Location/Qualifiers
 1. 832
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5416818"
 /clone_lib="NIH_MGC_87"
 /tissue_type="mammary adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.383 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC library."
 231 a 205 c 186 g 208 t 2 others

FEATURES
source

Query Match 19.7%; Score 701; DB 10; Length 832;
 Best Local Similarity 97.5%; Pred. No. 3.1e-164;
 Matches 785; Conservative 0; Mismatches 12; Indels 8; Gaps 7;

QY 2405 cagtcacacagcaggtgaaattctgagtcagcgaattccacacctaaacacc 2464
 DB 29 CAGGTATCACCAGGTGAATCTGGAAATTTCTAGCTGAGCAATCCACCTTAACACC 88

QY 2465 agttcccccagcaacagcagcgcgaaggatgtgccatacaaaagaactcttaagcagctga 2524
 DB 89 AGTTCCTCCAGCAACAGCAGCAGCAAGGATGTGCCATACAAAGAACTCTTAAAGCAGCTGA 148

QY 2525 ctcccccagcagcagcagctcttattgacctggcggcaacctcttctcacacccatact 2584
 DB 149 CTTCCTCAGCAGCATGGATCTTTA-TGACCTTGGGGGAACCTTTCTTACACACCCATACT 207

QY 2585 tctctgccagcttgagcagcagcagctatcgcttttaacaacattttgaaggcctactcac 2644
 DB 208 TCTCTGCCAGCTTGGCAGCAGCAGCAGCTATCGCTTTTACAACTTTTGAAGGCGCTACTCAC 267

QY 2645 ttctagaccaggaagtgggatatgtcccaaggtctcagctttgtagcaggcattttgcttc 2704
 DB 268 TTCTAGACCAGGAAGTGGGATATGCCAAAGTCTCAGCTTTGTAGCAGGCATTTGCTTC 327

QY 2705 ttcatatgagtgaagaaggcgttttaaaatgctcaagtcttgatgtttgacatggggc 2764
 DB 328 TTCTATATGATGAGGAGGAGGCTTTAAATGCTCAAGTTTCTGATGTTGACATGGGGC 387

QY 2765 tgcgaaacagtcagtcggcagacatgattattttacagatccagatgtaccagctctcga 2824
 DB 388 TCGGAAACAGTATCGGCCAGACATGATTTTACAGATCCAGATGTACCAGCTCTCGA 447

QY 2825 gttgtcttcattattaccacagagacctctacaataccctggagagcagcagatcgccc 2884
 DB 448 GGTGTCTCATGATTACCAACAGAGACCTCTACAATCACCTGGAGGAGCAGAGATCGGCC 507

QY 2885 ccagctctacgctcccctggttctcaccatgcttgcctccagcttcccctgggat 2944
 DB 508 CCAGCCTCTACGCTGCCCTGGTTCCTCACCATGTTTGCCTTCACAGTTCCCGCTGGGAT 567

QY 2945 tcgtagccagagcttttgatagatttttcttcagggaacagaggtcatatttaa-agtg 3003
 DB 568 TCGTAGCCAGAGCTTTGATATGATTTTCTTCAGGGAACAGAGTTCATATTTAACAGTG 627

QY 3004 gctttaagtctgttgggaagccataagcccttgattctcagcagatgaataacaccc 3063

DB 628 GCTTTAAGTCTGNTGGGAAGCCATAAGCCCTTGATTTCTGCAGCATGAACACCTAGAAACC 687
 QY 3064 atagtgtacttat-aaaaagcagcctaccacaccttggc-ttggtacagat-ggaaaaag 3120
 DB 688 ATAGTGTGACTTTATATAAAAGCAGCGCTAACCAAGCTTGCTTTGTGTACAGATGGGAAAAG 747
 QY 3121 accatcaatcaggtgatttgaaatggacatcgctaaa--cagttacaagcttatgaagttg 3178
 DB 748 ACCATCAATCAGGTATTGGAATGGACATCGGTAAACAGTTTACAAGCTTATGAAGTTG 807

QY 3179 agtacc-acgtctctcaagaagaac 3202
 DB 808 AGTACCAACGCTCCTTCAAGAAGAAC 832

RESULT 13
 AL570425 851 bp mRNA linear EST 16-FEB-2001
 LOCUS AL570425 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0D1020YF02 5
 DEFINITION prime, mRNA sequence.
 ACCESSION AL570425
 VERSION AL570425.1 GI:12926720
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 851)
 AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
 source
 1. 851
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0D1020YF02"
 /clone_lib="LTI_NFL006_PL2"
 /tissue_type="placenta"
 /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed by
 Life Technologies. Contact : Feng Liang Life Technologies,
 a division of Invitrogen 9800 Medical Center Drive
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com"

BASE COUNT 240 a 214 c 220 g 174 t 3 others
 ORIGIN

Query Match 19.4%; Score 689.8; DB 9; Length 851;
 Best Local Similarity 98.9%; Pred. No. 2e-161;
 Matches 714; Conservative 1; Mismatches 4; Indels 3; Gaps 2;

QY 2836 gattaccacagagacctctacaat--cacctggagagcagcagatcgccccagcctct 2893
 DB 1 GATTACCACAGAGACTCTACAACTCACCCTTGGAGGAGCAGAGATCGGCCCGAGCTCT 60

QY 2894 acgtgccccctggtctccaccatgttgcctcagttcccgctgggattcgtagcca 2953
 DB 61 AGCGTGCCCCCTGGTCTCCTCACCATGTTTGCTTCACAGTTCCTCGGATTCGTAGCCA 120

QY 2954 gagtcttgatgatgtttttcttcagggaacagaggtcatatttaaagtggcttaagtc 3013
 DB 121 GAGTCTTGATGATGATTTTCTTCAGGGAACAGAGTTCATATTAAAGTGGCTTAAAGTC 180

```
QY 3014 ttttgggaagccataagcccttgattcgtcagcagtagaagcccttagaaccatagttgact 3073
|
|
|
Db 181 TGTTGGGAAGCAYTAAGCCCTTGATCTTCAGCAGTAGAAGAACTAGAAACCATAGTTGACT 240
|
|
|
QY 3074 ttataaaagcagcgtactccccacccctgggttggtacagatggaagaaacacatcaatcagg 3133
|
|
|
Db 241 TTATAAAAGCAGCGTACCCCAACCTTGCTTGGTACAGATGGAAGAACCATCAATCAGG 300
|
|
|
QY 3134 tatitgaatggacatcgctaaacagttacaagcttatgaagcttgatgaagcttccttc 3193
|
|
|
Db 301 TATTTGAATGGATCGCTAAGAGTTACAGCTTATGAAGTTGAGTACCCAGCTCCCTTC 360
|
|
|
QY 3194 aagaagaacttatgattccctctcctcctcagtgagcaacaaagaaatggataaattagaga 3253
|
|
|
Db 361 AAGAGAACTTATCGATTCCCTCTCTCTCAGTGACACCAAGAAATGGATAAATTAGAGA 420
|
|
|
QY 3254 aaacacacagcagcttaacgaaacagaaaccttgacctcttgaaacagttgagtgagaa 3313
|
|
|
Db 421 AAACCAACAGCAGCTTACGCAACAGAACCTTGACCTCCTTGAACAGTTGACAGGTGGCAA 480
|
|
|
QY 3314 atgtagagatcaagccttgagccaccattgagaagctcctcagcagtgagagcaagc 3373
|
|
|
Db 481 ATGGTAGATCCAAAGCCTTGAGGCCACCATTTGAGAGCTCCTTGAGCAGTGAGAGCAAGC 540
|
|
|
QY 3374 tgaagcagcgcatgcttaccttagaactggagcgttgccgctcgtcgcagcagtgagg 3433
|
|
|
Db 541 TGAAGCAGGCCATGCTTACCTTAGAAGTGGAGCGGTGGGCCCTGCTGCAGACGGTGGAGG 600
|
|
|
QY 3434 agctggcggcggagcagcagccagcagcggcggcggcggcggcggcggcggcggcggc 3493
|
|
|
Db 601 AGCTGGCGCGCGGA-SGCAGAGCCAGCGAGCGGAGCCTGAGTGACGACGCGCGGAGC 659
|
|
|
QY 3494 ccagggcgactcagcgtctcagggagagattgcaacacacatccacactgtccagggc 3553
|
|
|
Db 660 CCAGGGCGGCTGACGCTCTCAGGAGAGATTGCAACACATCCACACTGTCACGCGCC 719
|
|
|
QY 3554 tt 3555
|
|
|
Db 720 TT 721

RESULT 14
BE300005
LOCUS 60944594F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960783 5',
DEFINITION mRNA sequence.
ACCESSION BE300005
VERSION BE300005.1 GI:9183753
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mhc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov
Plate: L1CM54 row: k column: 24
High quality sequence start: 2
High quality sequence stop: 753.
Location/Qualifiers
1. .754
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2960783"

FEATURES
source
```

RESULT 15

```
/clone_lib="NIH_MGC_17"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: muscle; Vector: pORF7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```
BASE COUNT 227 a 186 c 177 g 164 t
ORIGIN
```

```
Query Match 19.2%; Score 683.4; DB 10; Length 754;
Best Local Similarity 99.1%; Pred. No. 7.5e-160;
Matches 750; Conservative 0; Mismatches 1; Indels 6; Gaps 6;
```

```
QY 1878 ctcagtgacacagacgacctcatgaacgaagaccttgaaatccaaagcaaacattct 1937
|
|
|
Db 1 CTCAGTGAGCACAGAGACGCTCATGAACGAAGGACTTTGAATCCAAAGCAAAACCATC- 59
|
|
|
QY 1938 tgggtattcgttggtgacctcctgtgaagaccggaggcattcctggaggcagcatatt 1997
|
|
|
Db 60 TGGTGAATCTGGTGGGACTCCTGTGAAGACCCGGGAGGCAATTCCTGGAGGCAGAGATAT 119
|
|
|
QY 1998 cctccgagtagccaccgccgagaggcgtgcgattcttccagcagatatgaagtatttc 2057
|
|
|
Db 120 CCTCCGAGTAGCCACCCCGCAGAGAGGCTGCGATTCTTCCAGCAGATATGAAGATTATTC 179
|
|
|
QY 2058 agagctgggagagcttccccacagctctcctttagaacacagcttctgaagctggccctt 2117
|
|
|
Db 180 AGAGCTGGGAGAGCTTCCCCACCATCTCCTTTAGAACCCAG-TTGTGAAGATGGGCC-T 237
|
|
|
QY 2118 tggcccccaccagaggaagaaagacatctcgtgagctccgagagctgtggcaaaa 2177
|
|
|
Db 238 TGGCCCCCACCAGAGGAAAGAAAGACATCTCGTGAGCTCCGAGAGCTGTGCAAAA 297
|
|
|
QY 2178 ggcattcttcaacagatattcgtcgttagaattggagaagaaaaaacacagctcccaagc 2237
|
|
|
Db 298 GGCATTATTTCAACAGATACTGCTGCTTAGAATGAGAAGGAAATCAGAAGCTCCAAGC 357
|
|
|
QY 2238 ctctgaaaaatgattcgtgaacagccctgaagctcgatttatgaagaaattactccctg 2297
|
|
|
Db 358 CTCTGAAAAATGA-TTGCTGAACAAGCCCTGAAAGCTCGATTATGAAAGAAATTTACTCCCTG 416
|
|
|
QY 2298 tcttaagaagtaactacagtggtgggaaaagatgcttagcactccaggaagatcaaaaa 2357
|
|
|
Db 417 TCTTAAAGAAAGTAACACTACAGTGTGGGAAAAGATGCTTAGCACATCCAGAGATCAAAAAT 476
|
|
|
QY 2358 taagttgacatggaaaaaatgcaactggcgtgttggtggcaaggtgtgccagctatcccg 2417
|
|
|
Db 477 TAAGTTTGACATGGAAAAAATGCACTCGGTGTGTGGCAAGGTGTGCCAGCTCATACCCG 536
|
|
|
QY 2418 aggtgaatctggaaatttctagctgagcaattccaccttaaacaccagcttccacagcaa 2477
|
|
|
Db 537 AGGTGAATCTGGAAATTTCTAGCTGAGCAATTCACCTTAACACACAGTTTCCACAGCAA 596
|
|
|
QY 2478 acagcagccaaagagtgatgcatacaagaactcttaagaagctgacttcccagcagca 2537
|
|
|
Db 597 ACAGCAGCCAAAGGATGTGCATACAAAGAACTCTTAAAGCAGCTGACTTCCCAAGCAGCA 656
|
|
|
QY 2538 tggattcttatgaccttggggaaacctttctacacacccataacttctctgcccagct 2597
|
|
|
Db 657 TGGGATTTTGA-TGACCTTGGGGAACTTTCCTTACACACCCATCTCTCTGCCCCAGCT 715
|
|
|
QY 2598 tggagcaggagcagctatcgctttacaacatttgaag 2634
|
|
|
Db 716 TGGAGCAGGACAGCTATCCG-TTACAACATTTTGAGG 751
|
|
|
```

BE561889
LOCUS 601346094F1 NIH_MGC_8 892 bp mRNA linear EST 15-AUG-2000
DEFINITION 601346094F1 NIH_MGC_8 Homo sapiens CDNA clone IMAGE:3678887 5',
mRNA sequence.
ACCESSION BE561889
VERSION BE561889.1 GI:9805609
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 892)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LCM356 row: 11 column: 24
High quality sequence start: 50
High quality sequence stop: 758.
Location/Qualifiers
1..892
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3678887"
/clone.lib="NIH_MGC_8"
/tissue.type="Burkitt lymphoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GCGACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 224 a 240 c 242 g 186 t
ORIGIN

Query Match 19.2%; Score 682.6; DB 10; Length 892;
Best Local Similarity 95.3%; Pred. No. 1.3e-159;
Matches 790; Conservative 0; Mismatches 29; Indels 10; Gaps 8;

Qy 1628 aaagaccatctgtgtgaaaggagccttgcctcagtcgagagtcctccatctctgagagctcctttaagctcc 1687
|||||
Db 63 AAGAGCCATCTGTGTGTAAGAGGAGG-CTTGCCCATCTCTGAGAGCTCTCTTTAAGCTCC 121
|||||

Qy 1688 tcggctccctcggagagcctcctcagtcgagagtcctccatctcccaagaagcagctc 1747
|||||
Db 122 TCGGCTCTCGGAGGACCTGTCCAGTACTCGGAGAGTCATCTCCAGAGAGCCAGCTC 181
|||||

Qy 1748 cgctgtcgccccagcaggccttcaggaggcgagcaaacaccctgagtcactt-ccccatc 1806
|||||
Db 182 CGCTGTGCGCCCGCAGAGGCGCTTCAGGAGCGAGCAACACCCTGAGTCACTTCCGCCCATC 241
|||||

Qy 1807 gaatgccagaacctccacaacctcccggggtcccggggttttcgcaaggaactt 1866
|||||
Db 242 GAATGCCAGGAACCTCCCAACCTGCCCCGGGGGTCCCCGGGGGTTCGCAAGAGAACTT 301
|||||

Qy 1867 atgaggtatcactcagtcagcagcagacgcctcatgaacgaaggactttgaatccaaa 1926
|||||
Db 302 ATGAGGTATCATCTAGTGACAGACAGAGCGCTCATGACGAAGAGGACTTTGATCCCAA 361
|||||

Qy 1927 gcaaacattcttggtgattctggtgggactcctgtgaagaccgccgagggattcctcgagg 1986
|||||
Db 362 GCAAAATTATCTTGGTGATTCTGTGGGACTCTCTGTGAAGACCCGGAGGCATTCCTCGAGG 421
|||||

Qy 1987 cagcagatatctctccagtagccaccgccgcagaaagcgctgcgattcttccagcagatat 2046
|||||
Db 422 CAGCAGATATTCTCTCCGAGTAGCCACCCTCCGAGAGGGGTGGATTTCTTCCAGCAGATAT 481
|||||

Qy 2047 gaagattattcagagctgggagagcttcccccaagatcctcttttagaaccagtttgtgaa 2106
|||||
Db 482 GAAGATTATTAGAGCTGGGAGAGCTTCCCGCAGCATCTCCTTTAGAACCACTTTGTGAA 541
|||||

Qy 2107 gatgggccccttggccccccaccagagaaagaaagacatctcgtgagctcggagag 2166
|||||
Db 542 GATGGGCCCTTTGGCCCCCCCACAGAGAGAAAGAGCATCTCGTGAGCTCCGAGAG 601
|||||

Qy 2167 ctgtggcaaaaaggctattcttcaacagatactgctgcttagaagtgagaagaaatcag 2226
|||||
Db 602 CTGTGGCAAAAGGCTATTCTTCAACAGATACTGCTGCTTAGAATGGCGAAGAAATCAG 661
|||||

Qy 2227 aagctcccaagcctctgaaaaatgatttgctgaacagcgcctggaagctcgattatgaagaa 2286
|||||
Db 662 AAGCTCCAAGCCTCTGAAAATGATTGTGTAACAAGCGCCTGAAGCTCGATTATGAAGAA 721
|||||

Qy 2287 attactccctg-tcttaagaagtaactacagttgtgga--aaagatgcttagcactcca 2343
|||||

Db 722 ATTACTCCCTGTCTTAAAGAAAGTAACACAGTGTGGGACCAAGATGGCTTAGCACTCCA 781
|||||

Qy 2344 ggaagatcaaaaaattaaqtttgacatggaaaaaatgcactcggtctgttgggcaaggtgtg 2403
|||||

Db 782 GGAAGATC-CCCATTTAGTTTGGCCTGTGTGACAATGC-CTCGGGGTGTTGGGCCAGGGGG 839
|||||

Qy 2404 ccagctcatcccgaggtgaaatctggaatttctagctgagcaattcc 2452
|||||

Db 840 CC-CGTGTCATCCCGCGCTGAACCTGGG--AGTTTTAGTGGGCAATTC 885
|||||

Search completed: September 9, 2002, 16:41:28
Job time: 9256 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: September 9, 2002, 19:18:41 : Search time 706.85 Seconds
(without alignments)
8634.973 Million cell updates/sec

Title: US-09-762-311-4_COPY_176_3730
Perfect score: 3555
Sequence: 1 atggaaccaataacattcac.....tccacactgtccagccctt 3555

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Description
Result No.	Score	Query Match	Length DB ID	
1	3555	100.0	3983 21	AAZ50906 Human TBC-1 cDNA f
2	3555	100.0	3988 21	AAZ50907 Human TBC-1 cDNA f
3	3026.4	85.1	3326 22	AAI57854 Human polynucleoti
4	2940.2	82.7	3727 21	AACT7404 Human OREF2959
5	2833.8	79.7	3281 22	AAI59640 Human polynucleoti
6	2510.4	70.6	4039 19	AAV05886 Human tbc-1 gene.
7	2112	59.4	2362 22	AAH18401 Human cDNA sequenc
8	1000.8	28.2	1422 22	AAI34787 cDNA encoding nove
9	999.4	28.1	2989 22	AAK94796 Human full-length

10	943.2	26.5	1269	22	AAI34785	cDNA encoding nove
11	785	22.1	883	22	AAH04450	Human CDNA clone (
12	574.2	16.2	849	22	AAK92006	Human CDNA 5'-end
13	574.2	16.2	849	22	AAK93666	Human TBC-1 partia
14	422.6	11.9	17590	21	AAZ50904	Human gene express
15	362	10.2	696	20	AAI15229	Drosophila melanog
16	306.2	8.6	5472	23	ABL02013	Human TBC-1 partia
17	284.4	8.0	99960	21	AAZ50905	Human secreted pro
18	280	7.9	458	20	AAI40323	Drosophila melanog
19	245.6	6.9	8105	23	ABL02012	Human CDNA clone (
20	228.4	6.4	555	22	AAH09958	Human RECAP polynu
21	132.8	3.7	3691	22	AAI59516	Human polynucleoti
22	132.8	3.7	4827	22	AAI59592	Human polynucleoti
23	132.8	3.7	5294	22	AAI57806	Drosophila melanog
24	128.8	3.6	3585	23	ABL12337	CDNA #1312 encodin
25	123.8	3.5	154	23	AAI58636	Drosophila melanog
26	119.4	3.4	2626	23	ABL18133	Drosophila melanog
27	119.4	3.4	7829	23	ABL18132	Drosophila melanog
28	108	3.0	2424	23	AAI71793	DNA encoding novel
29	83.8	2.4	1601	21	AAI33401	Arabidopsis thalia
30	79.6	2.2	6765	23	ABL12336	Drosophila melanog
31	78.4	2.2	300	20	AAI13067	Human gene express
32	75.2	2.1	1492	21	AAI38750	Arabidopsis thalia
33	73.8	2.1	1140	21	AAI97020	Human secreted pro
34	68.4	1.9	3342	22	AAI32789	Human secreted pro
35	68	1.9	2492	21	AAI78199	Human cancer assoc
36	65.2	1.8	1831	22	AAI64964	Human secreted pro
37	65	1.8	1117	22	AAI61302	Human polynucleoti
38	65	1.8	2035	22	AAI59600	Human cell cycle a
39	65	1.8	2482	22	AAI59516	Human polynucleoti
40	65	1.8	2666	22	AAH17866	Human cDNA sequenc
41	62.6	1.8	2531	23	ABL07359	Drosophila melanog
42	62.6	1.8	4661	23	ABL07358	Drosophila melanog
43	61.6	1.7	2801	22	AAI32643	Human cDNA encodin
44	61.6	1.7	2803	22	AAI27080	cDNA encoding nove
45	61.6	1.7	2903	22	AAI51641	Human polynucleoti

ALIGNMENTS

RESULT 1

AAZ50906

ID AAZ50906 standard; cDNA; 3983 BP.

XX AAZ50906;

AC AAZ50906;

XX 31-MAY-2000 (first entry)

XX Human TBC-1 cDNA from first transcript.

XX TBC-1; human; biallelic marker; chromosome 4; cell cycle regulator; SNP;
XX Single nucleotide polymorphism; tissue differentiation; prostate cancer;
XX linkage analysis; genetic map; detection; diagnosis; genotyping;
XX transgenic animal; screening; alternative splicing; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT 5'UTR 1..170

FT /tag= a

FT /note= "Spans through exon 1 and part of exon 2"

FT CDS 171..3677

FT /tag= b

FT /product= "Human TBC-1 protein"

FT 3'UTR 3736..3983

FT /tag= c

FT polyA_signal 3942..3947

FT /tag= d

XX WO200008209-A2.

Db 1731 ctgcaggaaactccatcagctggatctggatctcctctgctctgtagtaac 1790
Qy 1621 accagaaagagccatctgtctgtgaaagaggcccttgcctatctctgagatccttt 1680
Db 1791 accagaaagagccatctgtctgtgaaagaggcccttgcctatctctgagatccttt 1850
Qy 1681 aagctccctcgctctcggaggaacctctccagtaactcggagagatcatctccagaagag 1740
Db 1851 aagctccctcgctctcggaggaacctctccagtaactcggagagatcatctccagaagag 1910
Qy 1741 ccagctccgctgtgcgccagcaggccttcaggaggaggagcaaacacccctgagtcacttc 1800
Db 1911 ccagctccgctgtgcgccagcaggccttcaggaggaggagcaaacacccctgagtcacttc 1970
Qy 1801 cccatcgaaatccaggaacctctccaaacctcccgccgggggtcccggggtttccaaagg 1860
Db 1971 cccatcgaaatccaggaacctctccaaacctcccgccgggggtcccggggtttccaaagg 2030
Qy 1861 aaacttaaggtatcactcagtagacagacagacgcctcatgaacgaagagacttgaa 1920
Db 2031 aaacttaaggtatcactcagtagacagacagacgcctcatgaacgaagagacttgaa 2090
Qy 1921 tccaaagcaaacactcttggctgattctggtggactcctgtgaagacccggagcattcc 1980
Db 2091 tccaaagcaaacactcttggctgattctggtggactcctgtgaagacccggagcattcc 2150
Qy 1981 tggaggcagcgatattctctccgagtagccaccccgagaaggcgtgcgattcttcacgc 2040
Db 2151 tggaggcagcgatattctctccgagtagccaccccgagaaggcgtgcgattcttcacgc 2210
Qy 2041 agatatgaagattatcagagctggagagcttcccccagatctcttttagaacaggtt 2100
Db 2211 agaatagaagattatcagagctggagagcttcccccagatctcttttagaacaggtt 2270
Qy 2101 tgtgaagatggcccttggcccccacagagagaaagaaagacatctcgtgagctc 2160
Db 2271 tgtgaagatggcccttggcccccacagagagaaagaaagacatctcgtgagctc 2330
Qy 2161 cgagagctgtggcaaaaggctattcttcaacagatctgctgttagaattggagaagaa 2220
Db 2331 cgagagctgtggcaaaaggctattcttcaacagatctgctgttagaattggagaagaa 2390
Qy 2221 aatcagaagctccagcctctgaaaatgatttctgcaacagcgcctgaaagctcgattat 2280
Db 2391 aatcagaagctccagcctctgaaaatgatttctgcaacagcgcctgaaagctcgattat 2450
Qy 2281 gaagaaattactcctctcttaagaaagtaactacagtgctgggaaaaagatgcttagcact 2340
Db 2451 gaagaaattactcctctcttaagaaagtaactacagtgctgggaaaaagatgcttagcact 2510
Qy 2341 ccaggaagatcaaaaattaaatttgacatggaaaaaatgcaactcggctgttgggcaaggt 2400
Db 2511 ccaggaagatcaaaaattaaatttgacatggaaaaaatgcaactcggctgttgggcaaggt 2570
Qy 2401 gtgcacgtctatccagagtgaaatctggaattcttagctgagcaattccacacttaaa 2460
Db 2571 gtgcacgtctatccagagtgaaatctggaattcttagctgagcaattccacacttaaa 2630
Qy 2461 caccaggtttccagcaacagcagccaaagagtgctgcaatacaagaactcttaagcag 2520
Db 2631 caccaggtttccagcaacagcagccaaagagtgctgcaatacaagaactcttaagcag 2690
Qy 2521 ctgacttcccgagcagatgcatattcttattgaccttggggcaaaccttctctacaccca 2580
Db 2691 ctgacttcccgagcagatgcatattcttattgaccttggggcaaaccttctctacaccca 2750
Qy 2581 tacttctctccagcttggagcagcagcagctatcgcttttacaacattttgaagcctac 2640
Db 2751 tacttctctccagcttggagcagcagcagcagctatcgcttttacaacattttgaagcctac 2810
Qy 2641 tcaacttctagaccaggaagtgggatattggcaaggtctcagcttcttagcaggcattttg 2700
Db 2811 tcaacttctagaccaggaagtgggatattggcaaggtctcagcttcttagcaggcattttg 2870

Qy 2701 cttcttcataatgagtgagggaaggcgttttaaaatgctcaagttctctgattttgacatg 2760
Db 2871 cttcttcataatgagtgagggaaggcgttttaaaatgctcaagttctctgattttgacatg 2930
Qy 2761 gggctcgggaaaaacagtagtcggccagacatgatttttacagataccagatataccagctc 2820
Db 2931 gggctcgggaaaaacagtagtcggccagacatgatttttacagataccagatataccagctc 2990
Qy 2821 tcgagggttcttcattattaccacagagacctctacaatcacctcggaggagacagatc 2880
Db 2991 tcgagggttcttcattattaccacagagacctctacaatcacctcggaggagacagatc 3050
Qy 2981 ggcceagcctctacgtgccccctggttctcaacctgtttgctcacaagttcccgctg 2940
Db 3051 ggcceagcctctacgtgccccctggttctcaacctgtttgctcacaagttcccgctg 3110
Qy 2941 ggaatcgtagcagagctctttgatattttcttcagggaacagaggtcatatttaaa 3000
Db 3111 ggaatcgtagcagagctctttgatattttcttcagggaacagaggtcatatttaaa 3170
Qy 3001 gtggtcttaagctctgttgggaagccataagcccttgattctctgcagcatgaaacactagaa 3060
Db 3171 gtggtcttaagctctgttgggaagccataagcccttgattctctgcagcatgaaacactagaa 3230
Qy 3061 acctagttgaacttataaaaagcacgcctacccaaccttggcttgggtacagatgaaaag 3120
Db 3231 acctagttgaacttataaaaagcacgcctacccaaccttggcttgggtacagatgaaaag 3290
Qy 3121 acctaatcaggtatttgaatggacatcgctaaacagttacacagctttagaagttgag 3180
Db 3291 acctaatcaggtatttgaatggacatcgctaaacagttacacagctttagaagttgag 3350
Qy 3181 taccacgtccttcaagaagaacttatcgattctctctctctcagtgacaacccaagaagt 3240
Db 3351 taccacgtccttcaagaagaacttatcgattctctctctcagtgacaacccaagaagt 3410
Qy 3241 gataaattagagaacaaacacagcagcttacgcaaacacagaaaccttgacctcttgaaacag 3300
Db 3411 gataaattagagaacaaacacagcagcttacgcaaacacagaaaccttgacctcttgaaacag 3470
Qy 3301 ttgaggtggcaaatggttagatccaaagccttgaggccaccattgagaagctcctgagc 3360
Db 3471 ttgaggtggcaaatggttagatccaaagccttgaggccaccattgagaagctcctgagc 3530
Qy 3361 agtgagagcagctgaagcggccatgcttaccttagaactggagcggctggccctgctg 3420
Db 3531 agtgagagcagctgaagcggccatgcttaccttagaactggagcggctggccctgctg 3590
Qy 3421 cagacgttgagagagctcgggcgagcagcagcagcagcagcagcagcagcagcagcagcagc 3480
Db 3591 cagacgttgagagagctcgggcgagcagcagcagcagcagcagcagcagcagcagcagcagc 3650
Qy 3481 acgagcccgagcccgagcggcgactgacagctctgcaggagagattgcaaacaccatccca 3540
Db 3651 acgagcccgagcccgagcggcgactgacagctctgcaggagagattgcaaacaccatccca 3710
Qy 3541 caactgtccaggcctt 3555
Db 3711 caactgtccaggcctt 3725

RESULT 2

AAZ50907 ID AAZ50907 standard; cDNA; 3988 BP.

XX AAZ50907;

XX AC AAZ50907;

XX DT 31-MAY-2000 (first entry)

XX Human TBC-1 cDNA from second transcript.

XX TBC-1; human; biallelic marker; chromosome 4; cell cycle regulator; SNP;

Db 1376 aggatagagggaatgaattcttccaaaacaaactagaaactgcaaaaggcacctgacagaca 1435
Qy 1261 ttaaccaatcaggagcaggagacactattttgaaagaggttcagaaattgagaccagagaat 1320
Db 1436 ttaaccaatcaggagcaggagacactattttgaaagaggttcagaaattgagaccagagaat 1495
Qy 1321 gagcagcagagagaatgaattgattattctttcttcttgagatgttttatatgaagagaaacag 1380
Db 1496 gagcagcagagagaatgaattgattattctttcttcttgagatgttttatatgaagagaaacag 1555
Qy 1381 aagaacacacatccatattgaggagatgaagcagacatcgcagatggcagcagagaaatatt 1440
Db 1556 aagaacacacatccatattgaggagatgaagcagacatcgcagatggcagcagagaaatatt 1615
Qy 1441 ggaagtgaattaccacccagtgccactcgatttaggttagatatactgctgaaacaaacagca 1500
Db 1616 ggaagtgaattaccacccagtgccactcgatttaggttagatatactgctgaaacaaacagca 1675
Qy 1501 aagagatctttaacagagctctttagaagatattttgtcccggtgtaataaagccagaggc 1560
Db 1676 aagagatctttaacagagctctttagaagatattttgtcccggtgtaataaagccagaggc 1735
Qy 1561 ctgcaggaaacactccatcagctggatctggtatagctccctgtctagtagtaccattaaagtaac 1620
Db 1736 ctgcaggaaacactccatcagctggatctggtatagctccctgtctagtagtaccattaaagtaac 1795
Qy 1621 accagcaaaagccatctgtgtgtgaaagaggccttgccactctctgtagagctccttt 1680
Db 1796 accagcaaaagccatctgtgtgtgaaagaggccttgccactctctgtagagctccttt 1855
Qy 1681 aagctctcgctctcgaggaaacctgccaagtccagtgactcggagagctatctccagagaag 1740
Db 1856 aagctctcgctctcgaggaaacctgccaagtccagtgactcggagagctatctccagagaag 1915
Qy 1741 ccagctccgctgtgcgccagcaggccttcaggaggcgagcaaacacctgagtcacttc 1800
Db 1916 ccagctccgctgtgcgccagcaggccttcaggaggcgagcaaacacctgagtcacttc 1975
Qy 1801 cccatgaatgcaggaaacctccacaacctgcgcgggggtcccccgggggttttcgcaagg 1860
Db 1976 cccatgaatgcaggaaacctccacaacctgcgcgggggtcccccgggggttttcgcaagg 2035
Qy 1861 aaacttatgaggtatcactcagtgagcacagagacgctcatgaacaaaggactttgaa 1920
Db 2036 aaacttatgaggtatcactcagtgagcacagagacgctcatgaacaaaggactttgaa 2095
Qy 1921 tccaaagcaaacactcttgggtgattctgtgtgggactcctgtgaagaccggaggcattcc 1980
Db 2096 tccaaagcaaacactcttgggtgattctgtgtgggactcctgtgaagaccggaggcattcc 2155
Qy 1981 tggaggcagcagatattctcccgagtagtagccaccgccgagagggcggtatcttccagc 2040
Db 2156 tggaggcagcagatattctcccgagtagtagccaccgccgagagggcggtatcttccagc 2215
Qy 2041 agatataagattattcagagctgggagagcttcccccacgactcctctttagaacaggtt 2100
Db 2216 agatataagattattcagagctgggagagcttcccccacgactcctctttagaacaggtt 2275
Qy 2101 tgtgaaagtgggccccttggccccccaccagagagaaagaaagacatctcgtgagctc 2160
Db 2276 tgtgaaagtgggccccttggccccccaccagagagaaagaaagacatctcgtgagctc 2335
Qy 2161 cgagagctgtggcaaaaggctattcttcaacagatactgctgtctagaaaggagaaagaa 2220
Db 2336 cgagagctgtggcaaaaggctattcttcaacagatactgctgtctagaaaggagaaagaa 2395
Qy 2221 aatcagagctccagcctctgaaatgatttctgaaacagccttgaagctcgattat 2280
Db 2396 aatcagagctccagcctctgaaatgatttctgaaacagccttgaagctcgattat 2455
Qy 2281 gaagaaattactcctgtctttaaagaagtaactacagtggtgggaaaaagatgcttagcact 2340
Db 2456 gaagaaattactcctgtctttaaagaagtaactacagtggtgggaaaaagatgcttagcact 2515

Qy 2341 ccaggagagatcaaaaaataagatttgacatggaaaaataatgcactcggctgttgggcaagggt 2400
Db 2516 ccaggagagatcaaaaaataagatttgacatggaaaaataatgcactcggctgttgggcaagggt 2575
Qy 2401 gtgcccagctcatcacccagagtgaaatctggaattttctagctgagcaattccaccttaaa 2460
Db 2576 gtgcccagctcatcacccagagtgaaatctggaattttctagctgagcaattccaccttaaa 2635
Qy 2461 caccagtttcccagcaaacacagcagccaaaggatgcatatacaagaactctttaaagcag 2520
Db 2636 caccagtttcccagcaaacacagcagccaaaggatgcatatacaagaactctttaaagcag 2695
Qy 2521 ctgacttcccagcagcagctgcgattcttattgacotttgggggaaaccttctctacacacca 2580
Db 2696 ctgacttcccagcagcagctgcgattcttattgacotttgggggaaaccttctctacacacca 2755
Qy 2581 tactctctcccagcagctgcgagcagcagctatcgctttacaacattttgaagccctac 2640
Db 2756 tactctctcccagcagctgcgagcagcagctatcgctttacaacattttgaagccctac 2815
Qy 2641 tcacttctagaccaggaagtgggatatgtccaaggtctcagcttttagcagggcattttg 2700
Db 2816 tcacttctagaccaggaagtgggatatgtccaaggtctcagcttttagcagggcattttg 2875
Qy 2701 cttcttcaatagctgagggaaggcgctttaaaatgctcaagttcttctgagtttgcacatg 2760
Db 2876 cttcttcaatagctgagggaaggcgctttaaaatgctcaagttcttctgagtttgcacatg 2935
Qy 2761 gggctgcgaaaaacagtagtcggccagacatgattattttacagatccagatgacagctc 2820
Db 2936 gggctgcgaaaaacagtagtcggccagacatgattattttacagatccagatgacagctc 2995
Qy 2821 tcgaggttgcttcatgattaccacagagacctctacaatcacctggaggagcagagatc 2880
Db 2996 tcgaggttgcttcatgattaccacagagacctctacaatcacctggaggagcagagatc 3055
Qy 2881 ggccccagcctctacgtgcctccctgggttctccacctggttgcctcacagttcccgctg 2940
Db 3056 ggccccagcctctacgtgcctccctgggttctccacctggttgcctcacagttcccgctg 3115
Qy 2941 ggattctgacagagctcttctgatatgattttcttcagggaacagaggttcataattaaa 3000
Db 3116 ggattctgacagagctcttctgatatgattttcttcagggaacagaggttcataattaaa 3175
Qy 3001 gtggctttaagctctgttgggaagccataagcccttgattctgcagcatgaaacctagaa 3060
Db 3176 gtggctttaagctctgttgggaagccataagcccttgattctgcagcatgaaacctagaa 3235
Qy 3061 accatagttgactttataaaaagcacgctaccacacctggcttgggttacagatggaaaag 3120
Db 3236 accatagttgactttataaaaagcacgctaccacacctggcttgggttacagatggaaaag 3295
Qy 3121 accatcaatcaggtattttgaaatggacatcgctaaacagttacaaagcttataagagttgag 3180
Db 3296 accatcaatcaggtattttgaaatggacatcgctaaacagttacaaagcttataagagttgag 3355
Qy 3181 tacacgcttccaaagaagaacttatcgattctctctctcagtgacaaaccaagaagt 3240
Db 3356 tacacgcttccaaagaagaacttatcgattctctctctcagtgacaaaccaagaagt 3415
Qy 3241 gataaattagaaaaaaccaacagcagcttacgaaaaacagaccttgacctcttgaaacag 3300
Db 3416 gataaattagaaaaaaccaacagcagcttacgaaaaacagaccttgacctcttgaaacag 3475
Qy 3301 ttgcaggttgccaaatggttaggataccaaagccttgaggccaccattgagaagctcctgagc 3360
Db 3476 ttgcaggttgccaaatggttaggataccaaagccttgaggccaccattgagaagctcctgagc 3535
Qy 3361 agtgagagcagctgaagcagggccatctctacotttagaactggagcgggtcgccctctg 3420
Db 3536 agtgagagcagctgaagcagggccatctctacotttagaactggagcgggtcgccctctg 3595

Db 943 cagatggcagcagagcaatattggaaagtgaattaccaccacagtcgccactcgatttaggcta 1002
QY 1480 gatatgtgaaaaaacaagcaagagatctttaacagagtcctttagaagaagtattttgtcc 1539
Db 1003 gatatgtgaaaaaacaagcaagagatctttaacagagtcctttagaagaagtattttgtcc 1062
QY 1540 cgggttaataaagccagaggcctgcaggaacactccatcagtgtagctgtagatgctcc 1599
Db 1063 cgggttaataaagccagaggcctgcaggaacactccatcagtgtagctgtagatgctcc 1122
QY 1600 ctgctagtacattaaagcaacacagcaagagccatctgtgtgaaagagagcccttg 1659
Db 1123 ctgctagtacattaaagcaacacagcaagagccatctgtgtgaaagagagcccttg 1182
QY 1660 cccatctctgagagtcctttaagctcctcgctcctcgaggacctgtccagtgactcg 1719
Db 1183 cccatctctgagagtcctttaagctcctcgctcctcgaggacctgtccagtgactcg 1242
QY 1720 gagagtcattcccaagaagccagctccgctgtgcgcccagcagcccttcaggaaggca 1779
Db 1243 gagagtcattcccaagaagccagctccgctgtgcgcccagcagcccttcaggaaggca 1302
QY 1780 gcaaacacctgagtcacttcccacgaatgccaggaacctcccaacctgcccgggg 1839
Db 1303 gcaaacacctgagtcacttcccacgaatgccaggaacctcccaacctgcccgggg 1362
QY 1840 tccccgggggttcgaaaggaaacttatgaggtatcactcagtgagcacagagacgct 1899
Db 1363 tccccgggggttcgaaaggaaacttatgaggtatcactcagtgagcacagagacgct 1422
QY 1900 catgaacgaaggaacttgaatccaaagcaaacactctggtgattctggtgggactcct 1959
Db 1423 catgaacgaaggaacttgaatccaaagcaaacactctggtgattctggtgggactcct 1482
QY 1960 gtgaagaccggaggaacttcctggaggcagcagatattccctccgagtagccaccgccag 2019
Db 1483 gtgaagaccggaggaacttcctggaggcagcagatattccctccgagtagccaccgccag 1542
QY 2020 aaggcgtgcgattcttcagcagatataagaattattcagagctgggagagcttccccca 2079
Db 1543 aaggcgtgcgattcttcagcagatataagaattattcagagctgggagagcttccccca 1602
QY 2080 cgatcctcttagaccagtttgtgaagatggccctttggccccccaccagagaaag 2139
Db 1603 cgatcctcttagaccagtttgtgaagatggccctttggccccccaccagagaaag 1662
QY 2140 aaaaagacatctcgtgagctcgagagctgtggcaaaaggctattcttcaacagatactg 2199
Db 1663 aaaaagacatctcgtgagctcgagagctgtggcaaaaggctattcttcaacagatactg 1722
QY 2200 ctgcttagaataaggaaagaaatacagaagctccaaagcctctgaaaaatgatttgtgaac 2259
Db 1723 ctgcttagaataaggaaagaaatacagaagctccaaagcctctgaaaaatgatttgtgaac 1782
QY 2260 aagcgcctgaagctcgattatgaagaaattactcctgtcttaaaagaagtaactacagtg 2319
Db 1783 aagcgcctgaagctcgattatgaagaaattactcctgtcttaaaagaagtaactacagtg 1842
QY 2320 tgggaaaaagatcttagcactccaggaagatcaaaaaatgaatttgacatggaaaaaatg 2379
Db 1843 tgggaaaaagatcttagcactccaggaagatcaaaaaatgaatttgacatggaaaaaatg 1902
QY 2380 cactcgctgttggcaaggtgtccacgtcatccagaggtgaattctggaaatttcta 2439
Db 1903 cactcgctgttggcaaggtgtccacgtcatccagaggtgaattctggaaatttcta 1962
QY 2440 gctgagcaattccaccttaaacacaggttccagcaaacagcagccaaagatgtgcca 2499
Db 1963 gctgagcaattccaccttaaacacaggttccagcaaacagcagccaaagatgtgcca 2022
QY 2500 tacaagaactcttaagcagctgacttcccagcagcatcgctattctttagccttggg 2559

Db 2023 tacaagaactcttaagcagctgacttcccagcagcatcgattcttattgaccttggg 2082
QY 2560 cgaaccttctacacacccatacttctctgtcccagcttgagcagcagctatcgctt 2619
Db 2083 cgaaccttctacacacccatacttctctgtcccagcttgagcagcagctatcgctt 2142
QY 2620 tacaacatttgaaggcctactcactcttagaccaggaagtgggatatatgccaaggtctc 2679
Db 2143 tacaacatttgaaggcctactcactcttagaccaggaagtgggatatatgccaaggtctc 2202
QY 2680 agctttgtagcaggcattttgtcttctcatatgagtgagggaagcgctttaaaaatgctc 2739
Db 2203 agctttgtagcaggcattttgtcttctcatatgagtgagggaagcgctttaaaaatgctc 2262
QY 2740 aagttcttgatgtttgacatgggctcggaacagtatcgccagcagacatgattatttta 2799
Db 2263 aagttcttgatgtttgacatgggctcggaacagtatcgccagcagacatgattatttta 2322
QY 2800 cagatccagatgtaccagctctcagaggtgtctcatgattaccacagagaccttacaat 2859
Db 2323 cagatccagatgtaccagctctcagaggtgtctcatgattaccacagagaccttacaat 2382
QY 2860 cacttgaggagcagcagatcgccccagcctctacgtgccccctgttctctaccatg 2919
Db 2383 cacttgaggagcagcagatcgccccagcctctacgtgccccctgttctctaccatg 2442
QY 2920 tttgctcacagttcccgctgggattctgagccagagcttttgatattttttctcag 2979
Db 2443 tttgctcacagttcccgctgggattctgagccagagcttttgatattttttctcag 2502
QY 2980 ggaacagaggtcataatttaaagtggtcttaagctctgttgggaagccataaagcccttgatt 3039
Db 2503 ggaacagaggtcataatttaaagtggtcttaagctctgttgggaagccataaagcccttgatt 2562
QY 3040 ctgcagcatgaaaaacctagaaccatagttgactttataaaaaagcacgtaccaccaactt 3099
Db 2563 ctgcagcatgaaaaacctagaaccatagttgactttataaaaaagcacgtaccaccaactt 2622
QY 3100 ggcctgtgacagatggaagaaccataatcaggtatttgaaatggacatcgctaaacag 3159
Db 2623 ggcctgtgacagatggaagaaccataatcaggtatttgaaatggacatcgctaaacag 2682
QY 3160 ttacaagcttatgaagttgagtcaccacgtctctcaagaagaacttctcgattctctctct 3219
Db 2683 ttacaagcttatgaagttgagtcaccacgtctctcaagaagaacttctcgattctctctct 2742
QY 3220 ctgcagtcacaacccaagaatggataaattagagaacacacacagcagcttaccgcaaacag 3279
Db 2743 ctgcagtcacaacccaagaatggataaattagagaacacacacagcagcttaccgcaaacag 2802
QY 3280 aacctgacctcttgaaacagttgcaggtggcaaatggtaggatccaaagccttgaggcc 3339
Db 2803 aacctgacctcttgaaacagttgcaggtggcaaatggtaggatccaaagccttgaggcc 2862
QY 3340 accattggaagctcctgagcagtgagagcaagctggaagcaggccatgcttaccctagaa 3399
Db 2863 accattggaagctcctgagcagtgagagcaagctggaagcaggccatgcttaccctagaa 2922
QY 3400 ctggagcaggtcggccctgctgcagcaggttgaggagctcgcgcgcgagcagcagccc 3459
Db 2923 ctggagcaggtcggccctgctgcagcaggttgaggagctcgcgcgcgagcagcagccc 2982
QY 3460 agcgaccgggagcctgagtcacgcagcccgagccacggcgagctgacagctctgcagg 3519
Db 2983 agcgaccgggagcctgagtcacgcagcccgagccacggcgagctgacagctctgcagg 3042
QY 3520 agagattgcaacacacatcccaactgtccaggcctt 3555
Db 3043 agagattgcaacacacatcccaactgtccaggcctt 3078

AACT77404	standard; cDNA; 3727 BP.
AACT77404	
08-FEB-2001	(first entry)
Human ORFX ORF2959	polynucleotide sequence SEQ ID NO:5917.
Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antichyroid; antianaemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; systemic malaria; autoimmune disorder; asthma allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive; ss.	
OS	Homio sapiens.
XX	
PN	W0200058473-A2.
XX	
PD	05-OCT-2000.
XX	
PF	31-MAR-2000; 2000WO-US08621.
XX	
PR	31-MAR-1999; 99US-0127607.
PR	02-APR-1999; 99US-0127636.
PR	05-APR-1999; 99US-0127728.
PR	30-MAR-2000; 2000US-0540763.
XX	
PA	(CURA-) CURAGEN CORP.
XX	
PI	Shmkets RA, Leach M;
XX	
DR	WPI; 2000-602362/57.
DR	P-PSDB; AAB43195.
XX	
PT	Novel nucleic acids and peptides derived from open reading frame X,
PT	useful for treating e.g. cancers, proliferative disorders,
PT	neurodegenerative disorders and cardiovascular disease -
XX	
PS	Claim 5; Page 5094-5096; 5507pp; English.
XX	
CC	AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC	which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC	sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC	antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC	osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC	immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC	antidiabetic; hypotensive; dermatological; immunosuppressive;
CC	antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC	antichyroid; and antianaemic. The sequences can be used for determining
CC	the presence of or predisposition to, or preventing or treating
CC	pathological conditions associated with an ORFX-associated disorder. The
CC	nucleic acids can be used to express ORFX proteins in gene therapy
CC	vectors. The proteins and nucleic acids may be used to treat cancers,
CC	proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC	graft vs host disease, cardiovascular disease, diabetes mellitus,
CC	hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC	erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC	bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC	allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC	nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC	coagulation; to inhibit thrombosis; and as a contraceptive.
XX	
SQ	Sequence 3727 BP; 1031 A; 889 C; 931 G; 875 T; 1 other;

Qy	1552	gccagagccctgcaggaaacctccatcagcttggtatctggatagctccctgtctagtaca	1611
Db	1024	gccagagccctgcaggaaacctccatcagcttggtatctggatagctccctgtctagtaca	1083
Qy	1612	ttaagtaaacacagcaagaagccatctgtgttgaaaaggaggccctgcccattctctgag	1671
Db	1084	ttaagtaaacacagcaagaagccatctgtgttgaaaaggaggccctgcccattctctgag	1143
Qy	1672	agctccttttaagctctctcgccctccctcgaggacctgtccagtgactcggagagtcacttc	1731
Db	1144	agctccttttaagctctctcgccctccctcgaggacctgtccagtgactcggagagtcacttc	1203
Qy	1732	ccagaagagccagctccgtctgtcgcccagcaggccttcaggagcgagcaaacacccgt	1791
Db	1204	ccagaagagccagctccgtctgtcgcccagcaggccttcaggagcgagcaaacacccgt	1263
Qy	1792	agtcacttcccactcgaatgccaggaacctccacaacctgcccgggggtcccgggggtt	1851
Db	1264	agtcacttcccactcgaatgccaggaacctccacaacctgcccgggggtcccgggggtt	1323
Qy	1852	tcgcaaaaggaaacttatgaggttatcactcagtgagcacagagacgcctcatgaacgaag	1911
Db	1324	tcgcaaaaggaaacttatgaggttatcactcagtgagcacagagacgcctcatgaacgaag	1383
Qy	1912	gactttgaatccaaaacaaacbatcttggtatctgtgtggactctctgtgaagaccgg	1971
Db	1384	gactttgaatccaaaacaaacbatcttggtatctgtgtggactctctgtgaagaccgg	1443
Qy	1972	aggcattccctggaggcagcagatattctcccgatagccaccgcagagaggcgtgcgat	2031
Db	1444	aggcattccctggaggcagcagatattctcccgatagccaccgcagagaggcgtgcgat	1503
Qy	2032	tcttcacgagatatgaagattattcagagctcgggagagcttcccacacgatctccttta	2091
Db	1504	tcttcacgagatatgaagattattcagagctcgggagagcttcccacacgatctccttta	1563
Qy	2092	gaaccagtttgtgaagatgggcctttggccccccaccagaggaaaaagagacatct	2151
Db	1564	gaaccagtttgtgaagatgggcctttgg - ccccccaccagaggaaaaagagacatct	1622
Qy	2152	cgtagctccgacagctgtggcaaaaggctattcttcacagatactgc - tgcctagaat	2210
Db	1623	cgtagctccgacagctgtggcaaaaggctattcttcacagatactgc - tgcctagaat	1682
Qy	2211	ggagaaggaaaaacagaagctccagcctctgaaatgatattgctgaacaagcgctgaa	2270
Db	1683	ggagaaggaaaaacagaagctccagcctctgaaatgatattgctgaacaagcgctgaa	1742
Qy	2271	gctcgattatgaagaaattactccctgtctttaaagaagtaactacagtgtgggaaaaagt	2330
Db	1743	gctcgattatgaagaaattactccctgtctttaaagaagtaactacagtgtgggaaaaagt	1802
Qy	2331	gcttagcactccaggaagatcaaaaattaagtttgacatgaaaaaatgcactgcgtgt	2390
Db	1803	gcttagcactcccgagggaatccaaaaattaagtttgacatgaaaaaatgcactgcgtgt	1862
Qy	2391	tgggcaaggtgtgcccagctcatccacgaggtgaaatctcgaaattcttagctgagcaatt	2450
Db	1863	tgggcaaggtgtgcccagctcatccacgaggtgaaatctcgaaattcttagctgagcaatt	1922
Qy	2451	ccaccttaaacaccagtttcccagcaaacagcagccaaagatgtgccatacaagaact	2510
Db	1923	ccaccttaaacaccagtttcccagcaaacagcagccaaagatgtgccatacaagaact	1982
Qy	2511	cttaaaacagctgaacttcccagcagcatgcgatttattgaccttgggcgaaaccttcc	2570
Db	1983	cttaaaacagctgaacttcccagcagcatgcgatttattgaccttgggcgaaaccttcc	2042
Qy	2571	tacacaccatacttcttgcgcagcttgagcaggaagcgtatcgctttacaacattt	2630
Db	2043	tacacaccatacttcttgcgcagcttgagcaggaagcgtatcgctttacaacattt	2102
Qy	2631	gaaggcttactcactctagacacaggaagtgggatttgccaaaggtctcagctttgtagc	2690

Db	2103	gaaggcctactcactctctagaccaggaagtgggattgccaaggtctcagctttagc	2162
Qy	2691	aggcattttgctcttcatactagtgagaaagcggttaaaatgctcaagttctgat	2750
Db	2163	agggcatttgcctcttcacatgagtgaggaagggcgtttaaaatgctcaagttctgat	2222
Qy	2751	gtttgacatggggctgcggaaacagatcggcgagacatgattatttacagatccagat	2810
Db	2223	gtttgacatggggctgcggaaacagatcggcgagacatgattatttacagatccagat	2282
Qy	2811	gtaccagctctcgagggttctctcatgattaccacagagacctctacaatcacctggagga	2870
Db	2283	gtaccagctctcgagggttctctcatgattaccacagagacctctacaatcacctggagga	2342
Qy	2871	gcaagagatcgggcccgagcctctacgtcgccctggttctccacacatgatttgcctcaca	2930
Db	2343	gcaagagatcgggcccgagcctctacgtcgccctggttctccacacatgatttgcctcaca	2402
Qy	2931	gtcccgctggagattogtagccagagctcttatgattttctccaagggaacagaggt	2990
Db	2403	gtcccgctggagattogtagccagagctcttatgattttctccaagggaacagaggt	2462
Qy	2991	catattaaagtggctttaagtctgttgggaagccataaagccctgattctgcagcatga	3050
Db	2463	catattaaagtggctttaagtctgttgggaagccataaagccctgattctgcagcatga	2522
Qy	3051	aaacctagaaaacctatgttgaatttataaaagcagcgtaccacaacctgtggtgtaca	3110
Db	2523	aaacctagaaaacctatgttgaatttataaaagcagcgtaccacaacctgtggtgtaca	2582
Qy	3111	gatgaaaaagaccatacaactcaggtatttgaattggacatcgctaacaagttacaagctta	3170
Db	2583	gatgaaaaagaccatacaactcaggtatttgaattggacatcgctaacaagttacaagctta	2642
Qy	3171	tgaagtgtagtcaccagctctctcaagaagaactatcgattctctctcctcagtgacaa	3230
Db	2643	tgaagtgtagtcaccagctctctcaagaagaactatcgattctctctcctcagtgacaa	2702
Qy	3231	ccaaagaaatggataaattagagaaaaaccaagcagcgttacgcaaacagaaaccttgacct	3290
Db	2703	ccaaagaaatggataaattagagaaaaaccaagcagcgttacgcaaacagaaaccttgacct	2762
Qy	3291	ccttgaaacagttgcaggtggcaaatggtaggtccaaagccttgagggccaccatttgagaa	3350
Db	2763	ccttgaaacagttgcaggtggcaaatggtaggtccaaagccttgagggccaccatttgagaa	2822
Qy	3351	gctcctgagcagtgagagcaagctgaagcaggccattgcttaccttagaactgagagcggtc	3410
Db	2823	gctcctgagcagtgagagcaagctgaagcaggccattgcttaccttagaactgagagcggtc	2882
Qy	3411	ggccctgctgcagacaggttgagagagctcgggcgggagcgagagccccagcaccggga	3470
Db	2883	g-- cctgcctgcagacaggttgagagagctcgggcgggagcgagagccccagcaccggga	2940
Qy	3471	gctcgagtcacgcagcccgagccccagggcgactgacagctctgcagagagattgtcaa	3530
Db	2941	gctcgagtcacgcagcccgagccccagggcgactgacagct--tgcaggagagattgtcaa	2999
Qy	3531	caccatcccacactgtccaggcctt	3555
Db	3000	caccat--caccactgtccaggcctt	3022

RESULT	5
AAI59640	
ID	AAI59640 standard; cDNA; 3281 BP.
XX	
XX	
AC	AAI59640;
XX	
DT	22-OCT-2001 (first entry)
XX	
DE	Human polynucleotide SEQ ID NO 3629.

XX	Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX	leukaemia; ss.
OS	Homo sapiens.
XX	
PN	WO200153312-A1.
XX	
PD	26-JUL-2001.
XX	
PF	26-DEC-2000; 2000WO-US34263.
XX	
PR	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-0552317.
PR	09-JUL-2000; 2000US-0598042.
PR	19-JUL-2000; 2000US-0620312.
PR	03-AUG-2000; 2000US-0653450.
PR	14-SEP-2000; 2000US-0662191.
PR	19-OCT-2000; 2000US-0693036.
PR	29-NOV-2000; 2000US-0727344.
XX	(HYSE-) HYSEQ INC.
PA	
PI	Tang YT, Liu C, Asundi V,
PI	Wang J, Wang Z, Wehrman T,
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX	
DR	WPI; 2001-442253/47.
DR	P-PSDB; AAM40484.
XX	
PT	Novel nucleic acids and polypeptides, useful for treating disorders
PT	such as central nervous system injuries -
XX	
PS	Claim 1; SEQ ID NO 3629; 10078pp; English.
XX	
CC	The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC	the encoded polypeptides (AAM38642-AAAM42213) with neurotropic.
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilisation of the activities such as: Immune system suppression,
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemias and
CC	C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
XX	
SQ	Sequence 3281 BP; 912 A; 812 C; 824 G; 733 T; 0 other;
Query Match 79.7%; Score 2833.8; DB 22; Length 3281;	
Best Local Similarity 98.3%; Pred. No. 0;	
Matches 2983; Conservative 0; Mismatches 32; Indels 20; Gaps 11	
Qy	541 gtgacgttgccgcacaagaagcctccgccggccctgatcgacgaagtgcatacgagaatttc 600
Dd	 1 gtagcgggtggcgcaagaaggctccgccggccctgatcgacgaagtgcatacgagaatttc 60
Qy	601 aatcacgtcacgcagccgggggtccgcagagcccccccacaccccccaccatgccgcg 660
Dd	 61 aatcacgtcacgcagccgggggtccgcagagcccccccacaccccccaccatgccgcg 120
Qy	661 cccacaggagccaggagacctgtgcgaggcccatgcgcaagtctcttccagaccgcgcg 720

QY 1782 aaacacccatgaagtcacatcccatcgaatgccaggaaccctccacaacactgcccggggtc 1841
|||||
DB 1261 aaacacccatgaagtcacatcccatcgaatgccaggaaccctccacaacactgcccggggtc 1320
QY 1842 cccgggggtttcgaaaggaaacttatgaggtatcaactcagtgagcagacagagcctca 1901
|||||
DB 1321 cccgggggtttcgaaaggaaacttatgaggtatcaactcagtgagcagagagcctca 1380
QY 1902 tgaacgaaggactttgaatccaaagcaaacctcttgggtattctctggtgggactcctgt 1961
|||||
DB 1381 tgaacgaaggactttgaatccaaagcaaacctcttgggtattctctggtgggactcctgt 1440
QY 1962 gaagaccggaggtatcctcgaggagcagcagatatctcctcgagtagccaccgcagaa 2021
|||||
DB 1441 gaagaccggaggtatcctcgaggagcagcagatatctcctcgagtagccaccgcagaa 1500
QY 2022 ggcgtcgattcttccagcagatatgaagattattcagagctgggagagcttcccccaag 2081
|||||
DB 1501 ggcgtcgattcttccagcagatatgaagattattcagagctgggagagcttcccccaag 1560
QY 2082 atctctttagaacacgtttgtgaagatgggccccttggccccccacacagaggaagaa 2141
|||||
DB 1561 atctctttagaacacgtttgtgaagatgggccccttggccccccacacagaggaagaa 1620
QY 2142 aagacatctcgtgagctccgagagctgtgcaaaagctattcttcaacagatactgct 2201
|||||
DB 1621 aagacatctcgtgagctccgagagctgtgcaaaagctattcttcaacagatactgct 1680
QY 2202 gcttagaattgagaggaataatcagaagctccaagctctgaaatgatttgcgtaacaa 2261
|||||
DB 1681 gcttagaattgagaggaataatcagaagctccaagctctgaaatgatttgcgtaacaa 1740
QY 2262 gcgcctgaactcgtattatgaagaaatactcctcgtcttaagaagataactacagtgct 2321
|||||
DB 1741 gcgcctgaactcgtattatgaagaaatactcctcgtcttaagaagataactacagtgct 1800
QY 2322 ggaagaatgcttagcactcaggaagatcaaaaataaattgacatggaataaaatgca 2381
|||||
DB 1801 ggaagaatgcttagcactcaggaagatcaaaaataaattgacatggaataaaatgca 1860
QY 2382 ctgcgctgttgggaaggtgtg-ccacgtcatccagaggtgaaatttggaattttctag 2440
|||||
DB 1861 ctgcgctgttgggaaggtgtgcccacgtcatccagaggtgaaatttggaattttctag 1920
QY 2441 ctgagcaattccacacttaaacacacagtttccacgcaaacagcagcgaagatgtgccat 2500
|||||
DB 1921 ctgagcaattccacacttaaacacacagtttccacgcaaacagcagcgaagatgtgccat 1980
QY 2501 acaagaactcttaagcagctgacttcccagcagcagcagcagcagcagcagcagcagc 2560
|||||
DB 1981 acaagaactcttaagcagctgacttcccagcagcagcagcagcagcagcagcagcagc 2040
QY 2561 gaaccttctctacacccatacttctctcctcagcagctggagcagcagcagcagcagcagc 2620
|||||
DB 2041 gaaccttctctacacccatacttctctcctcagcagctggagcagcagcagcagcagcagc 2100
QY 2621 acaacatttgaagcctactcacttcttagaccaggaagtgggatatggccaaggtctca 2680
|||||
DB 2101 acaacatttgaagcctactcacttcttagaccaggaagtgggatatggccaaggtctca 2160
QY 2681 gctttgtagcaggaatttgccttctcatatgagtgaggaagagcgcttttaaaatgctca 2740
|||||
DB 2161 gctttgtagcaggaatttgccttctcatatgagtgaggaagagcgcttttaaaatgctca 2220
QY 2741 agttcttgatgtttgacatggggtcggaacacagatcgccagacatgattattttac 2800
|||||
DB 2221 agttcttgatgtttgacatggggtcggaacacagatcgccagacatgattattttac 2280
QY 2801 agatccagatgtaccagctctcaggttgccttcatgattaccacagagacctctacaatc 2860
|||||
DB 2281 agatccagatgtaccagctctcaggttgccttcatgattaccacagagacctctacaatc 2340

QY 2861 acctggaggagcagcagatcgccccagcctctaagctgccccctgggtctcctaccatgt 2920
|||||
DB 2341 acctggaggagcagcagatcgccccagcctctaagctgccccctgggtctcctaccatgt 2400
QY 2921 ttgctcacagttcccgcgtgggattcgaccagagtccttttgatatatttttttcagg 2980
|||||
DB 2401 ttgctcacagttcccgcgtgggattcgaccagagtccttttgatatatttttttcagg 2460
QY 2981 gaacagaggtcataatttaaaagtggctttaagtctgttgggaagcacaataagcccttgattc 3040
|||||
DB 2461 gaacagaggtcataatttaaaagtggctttaagtctgttgggaagcacaataagcccttgattc 2520
QY 3041 tgacagcatgaaacctgaaagaccatagttagctttataaaagcacgctaccacaccttg 3100
|||||
DB 2521 tgacagcatgaaacctgaaagaccatagttagctttataaaagcacgctaccacaccttg 2580
QY 3101 gcttggtacagatgaaagaccatcaatcaggtatttgaattggacatcgctaaacagt 3160
|||||
DB 2581 gcttggtacagatgaaagaccatcaatcaggtatttgaattggacatcgctaaacagt 2640
QY 3161 tacaagcttatgaattgagtagcacagctccttcaagaagaacttatgattcctcctc 3220
|||||
DB 2641 tacaagcttatgaattgagtagcacagctccttcaagaagaacttatgattcctcctc 2700
QY 3221 tcagtgacaaccaaagaatggataaaattagagaaaccaaagcagcttagcgaacaga 3280
|||||
DB² 2701 tcagtgacaaccaaagaatggataaaattagagaaaccaaagcagcttagcgaacaga 2760
QY 3281 acctgacctccttgaacagttgaggtggcaaatggttaggatccaaagccttagggcca 3340
|||||
DB 2761 acctgacctccttgaacagttgaggtggcaaatggttaggatccaaagccttagggcca 2820
QY 3341 ccattgagaagcctcctgagcagtgagagcaagctgaagcagcagcagcagcagcagcagc 3400
|||||
DB 2821 ccattgagaagcctcctgagcagtgagagcaagctgaagcagcagcagcagcagcagcagc 2880
QY 3401 tgagcagctcgccctcgtcgcagcagctggaggtgagcggcgagcagcagcagcagcagc 3460
|||||
DB 2881 tgagcagctcgccctcgtcgcagcagctggaggtgagcggcgagcagcagcagcagcagc 2940
QY 3461 gcgaccgggagcctgagtgacgcagcagcagcagcagcagcagcagcagcagcagcagcagc 3520
DB 2941 gcgaccgggagcctgagtgacgcagcagcagcagcagcagcagcagcagcagcagcagcagc 3000
QY 3521 gagattgcaacacacatcccacactgtccaggcctt 3555
|||||
DB 3001 gagattgcaacacacatcccacactgtccaggcctt 3035

RESULT 6

AAV05886

ID AAV05886 standard; DNA; 4039 BP.

XX

AC AAV05886;

XX

DT 01-JUN-1998 (first entry)

XX

DE Human Tbc-1 gene.

XX

KW tre-2; BUB2; cdc16; mast cell; probe; hybridisation; subtraction method;
mouse; transcription factor; differentiation; proliferation; human; ds;
acute myelogenous leukaemia.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers
CDS 403..3828

FT /*tag= a

FT /product= Tbc1_protein

XX

PN US5700927-A.

XX

PD 23-DEC-1997.

XX 23-DEC-1994; 94US-0363300.
XX 23-DEC-1994; 94US-0363300.
XX (CHIL-) CHILDRENS MEDICAL CENT.
XX Richardson P, Zon L;
XX WPI; 1998-062437/06.
XX P-PSDB; AAW44777.
XX DNA encoding Tbc1 polypeptide - useful for treating leukaemia
XX Claim 3; Fig 1A-B; 22pp; English.
XX This is the nucleotide sequence which encodes a Tbc1 (tre-2, BUB2 and
XX cdc16) polypeptide, isolated from a bone marrow-derived mast cell DNA
XX library. The screen was carried out using a probe generated by a
XX subtraction method which compared mRNA expression in an undifferentiated
XX mast cell line P815 and cell line PC76 (P815 cells transformed to express
XX the murine GATA-1 transcription factor - a factor which controls the
XX expression of genes involved in mast cell differentiation). Tbc1 encodes
XX a protein involved in the coupling of cell proliferation to cell
XX differentiation, which can be used to treat leukaemia (especially acute
XX myelogenous leukaemia) by causing leukaemic cells to differentiate.
XX Sequence 4039 BP; 1042 A; 1096 C; 1060 G; 841 T; 0 other;
SQ

Query Match 70.6%; Score 2510.4; DB 19; Length 4039;
Best Local Similarity 83.7%; Pred. No. 0;
Matches 2949; Conservative 0; Mismatches 536; Indels 37; Gaps 8;

QY 1 atggaacaaataacattcacaggaagaaacatctgctt-ctaaagcgggtctcggtgga 59
DB 292 atggaggcaatacacattcacaggaagcaatccgtttccctaaagaaagtcctcgga 351
QY 60 ttttgccctgcagctggtt-ggctccctgcctgtgcatcctcctgaccacacatgccatg 117
DB 352 ctctcgctgcagctggttgggtctctctaccctgcatcctcctcaccactatgccatg 411
QY 118 ctgcccctgggttggctgaggtgcgaagactcagcagcagcagcagcagcagcagcagc 177
DB 412 ctgcctgggttgtagcaggttagaagactcagcagcagcagcagcagcagcagcagc 471
QY 178 gtaaccaagaagtcgggttgcgtttcaccctctgactgagatggaacctgagcca 237
DB 472 aggaacaaagcaagtcgccttgggttgcctcctggactcgggtgcgagcctgacctg 531
QY 238 gggagaagtcacagtggtatccctgactctattccagcatctttagtgcaagcctcag 297
DB 532 gggaaaagcaacattgggaccctgccaactgtctccagcatctttagtgcaagcctcag 591
QY 298 cgtgttcacaaactgattcacacagtcagaccacagttacttctgtctgtattgaag 357
DB 592 cgtgtgcacaaactgattcacacagtcagaccagcagcagcagcagcagcagcagc 651
QY 358 gtagcctgtctcacagcagagtatctgtctatgtttcaagcgcagatgatacaacaaa 417
DB 652 gaggatgctgcacagcagagcctctgtatgttttaagcagacgacgatacaacaaa 711
QY 418 gtgcctgagatcatcagctccatcctcagcggggaagatcgcgcgcgagagagctg 477
DB 712 gtgctgagatcatcagcctccatcctcggcggcgggaagatgcccgcgaggaagctg 771
QY 478 cactgcccctcgcagttcacacagcagcttttccaaagattcgaggtctctctcgccg 537
DB 772 cgttgcctcgcagttcacacagcagccttcgcacaaagtgcaggtctctctgaggc 831
QY 538 cgcgtacaggttggcgcacaaaggtctccgcggccctgactgacgagtgcatcgagaag 597
DB 832 cgggtgactgtggtcacaaagggccccaccgcgactgattgacgagtgatcagagaag 891

QY 598 ttcaatcacgtcagcgcagcgcgggtccgagagagccccccgcccccaacccgccatgcc 657
DB 892 ttcaaccatgtgagctgtggtcgcgaacgcgactgggaagcgcacccggggcagccatca 951
QY 658 gcgcacacagggagcagagagctgtgcgagggcccaatgcgcaagtctcttcccagccc 717
DB 952 gc-----gcctggccccagggcccatgcgcaaatccttctcacagcct 993
QY 718 ggcctgcctcgcctgctgaggaagagctgcagaggtggggggctccgagcagcagcggc 777
DB 994 ggaactgcctgctgctgctcaggaagaggttccaggagcgtagctccgcgagtgaacc 1053
QY 778 ttcttcagctccttcgagagagcagacattgagaacacacctattagcggagacacaatt 837
DB 1054 tt-----tagctcttggacaatgacatagagaacacctcatcgttggggcacaaatg 1107
QY 838 gtgcagccacagatcatcagagaaatcgaaactatgctcttccacgattggccagctgaa 897
DB 1108 gtgcagccacagacatggaggagaacgaaactatgcttccagatgtggccatctgaa 1167
QY 898 gttaccctcatcagctcctgacacacacacacacacacacacacacacacacacacac 957
DB 1168 gttaccctcatcagctcctgacacacacacacacacacacacacacacacacacacac 1227
QY 958 tcttttctcctcagggcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1017
DB 1228 tcttttctcctcagggcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1287
QY 1018 tc-----cggaggtggcgggttcttctgtctgttctgttctgttctgttctgttctgt 1071
DB 1288 tcgggtggcggcagtgggcgggttcttctgtctgttctgttctgttctgttctgttctgt 1347
QY 1072 gctcgggttgatgaataatgatgacctgaacacagccttccacggtgcccagtgccag 1131
DB 1348 gctcgggttgacgagatcatgatctgaagcaggttctcagcgttagctcggtgag 1407
QY 1132 cagacagctaaagcgcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1191
DB 1408 cagacgctaaagcgcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1467
QY 1192 ctctgtgagagatgagggaggaatgatttctccacacacacacacacacacacacacac 1251
DB 1468 ctctgtgaaagatgagggaggaatgatttctccacacacacacacacacacacacacac 1527
QY 1252 ctgcagcaatgaacaaatcagagcagcagcagcagcagcagcagcagcagcagcagc 1311
DB 1528 ttgacacacatgacaaatcagagcagcagcagcagcagcagcagcagcagcagcagc 1587
QY 1312 ccgaaatgagcagcagagaggaatgaattgatttcttctgttctgttctgttctgttctgt 1371
DB 1588 ccaagaacagcagcagagaggaatgaatttcttctgttctgttctgttctgttctgt 1647
QY 1372 ggaacacagaaagac 1431
DB 1648 gagaacacaaagac 1707
QY 1432 ggaatattggaagtgaattaccacacacacacacacacacacacacacacacacacac 1491
DB 1708 ggaatattggaagtgaattaccacacacacacacacacacacacacacacacacacac 1767
QY 1492 aacaagcaagagatctttaaagagctttaaagagctttaaagagctttaaagagctttaa 1551
DB 1768 aacagacaaagaggttcttaacagagcttcttaacagagcttcttaacagagcttct 1827
QY 1552 gccagagcctcagaaac 1611
DB 1828 gccagagcctcagaaac 1887
QY 1612 ttaagtaac 1671
DB 1888 ctaagtaac 1947

PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 18465; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-qt primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2362 BP; 657 A; 581 C; 578 G; 546 T; 0 other;

Query Match 59.4%; Score 2112; DB 22; Length 2362;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1444 agtgaattaccacccagtcgactcgatttagctagatgctgtaaaacaaagcaag 1503
Db 1 agtgaattaccacccagtcgactcgatttagctagatgctgtaaaacaaagcaag 60
Qy 1504 agatctttaacagagcttttagaagaattttgtcccggttaataaagcagggcgtg 1563
Db 61 agatctttaacagagcttttagaagaattttgtcccggttaataaagcagggcgtg 120
Qy 1564 caggaaacacccatcagtggtgactgtagctccctgtctagtaataagtaaacacc 1623
Db 121 caggaaacacccatcagtggtgactgtagctccctgtctagtaataagtaaacacc 180
Qy 1624 agcaaaagacatctgtgtgaaagaggccttgcccctctctgagagctcctttaag 1683
Db 181 agcaaaagacatctgtgtgaaagaggccttgcccctctctgagagctcctttaag 240
Qy 1684 ctctcgtcctcctcggaggaccgttcagtgactcggagagatcattccacagagacca 1743
Db 241 ctctcgtcctcctcggaggaccgttcagtgactcggagagatcattccacagagacca 300
Qy 1744 gctccgtgtgcgccagagccttcaggaggagcgaacacacccctgaactctcccc 1803
Db 301 gctccgtgtgcgccagagccttcaggaggagcgaacacacccctgaactctcccc 360

Qy 1804 atcgaatgcaggaacacctccacaaacctgcccgggggtccccgggggttttcgcaaaagaaa 1863
Db 361 atcgaatgcaggaacacctccacaaacctgcccgggggtccccgggggttttcgcaaaagaaa 420
Qy 1864 cttatgaggtatcactcactcagtgagcagagacgcctcctatgaacgaaggaactttgaatcc 1923
Db 421 cttatgaggtatcactcactcagtgagcagagacgcctcctatgaacgaaggaactttgaatcc 480
Qy 1924 aaagcaaaacatcttggtgattctggtgagactcctctgtaagaccgagagcattctctgg 1983
Db 481 aaagcaaaacatcttggtgattctggtgagactcctctgtaagaccgagagcattctctgg 540
Qy 1984 aggcagcagatattctccgagtagccaccccgcaagaaggcgtgcgattcttccagcaga 2043
Db 541 aggcagcagatattctccgagtagccaccccgcaagaaggcgtgcgattcttccagcaga 600
Qy 2044 tatgaagattattcagagctggagagcttcccccagcatctcctttagaaacagatttgt 2103
Db 601 tatgaagattattcagagctggagagcttcccccagcatctcctttagaaacagatttgt 660
Qy 2104 gaagatgggcccctttggccccccacccagaggaagaaagaaagagacatctcgtgagctccga 2163
Db 661 gaagatgggcccctttggccccccacccagaggaagaaagaaagagacatctcgtgagctccga 720
Qy 2164 gagctgtgcaaaagcctattcttcaacagatgactgctgttagaatggagaaggaat 2223
Db 721 gagctgtgcaaaagcctattcttcaacagatgactgctgttagaatggagaaggaat 780
Qy 2224 cagaagctcaagcctctgaaatgatttgcgaacaaagcgcctgaagctcgattatgaa 2283
Db 781 cagaagctcaagcctctgaaatgatttgcgaacaaagcgcctgaagctcgattatgaa 840
Qy 2284 gaattactcctcttcttaagaagtaactacagtgctgggaaagatgcttagcactcca 2343
Db 841 gaattactcctcttcttaagaagtaactacagtgctgggaaagatgcttagcactcca 900
Qy 2344 ggaagatcaaaaattgaatttgacatggaaaaaatgcactcggctgttggcgaaggtgtg 2403
Db 901 ggaagatcaaaaattgaatttgacatggaaaaaatgcactcggctgttggcgaaggtgtg 960
Qy 2404 ccagctcatcacccaggtgaaatctgaaatttctagctgagcaattcccacacctaaacac 2463
Db 961 ccagctcatcacccaggtgaaatctgaaatttctagctgagcaattcccacacctaaacac 1020
Qy 2464 cagtttcccgagcaaacagcagcacaagagtgatgccatacaaaagactcttaaaagcagctg 2523
Db 1021 cagtttcccgagcaaacagcagcacaagagtgatgccatacaaaagactcttaaaagcagctg 1080
Qy 2524 acttcccgagcagcatgogatttctattgaccttggcgaaaccttctcacacccatac 2583
Db 1081 acttcccgagcagcatgogatttctattgaccttggcgaaaccttctcacacccatac 1140
Qy 2584 ttctctgcccagcttgagcagcagcagctatcgctttacacatttttgaggcctactca 2643
Db 1141 ttctctgcccagcttgagcagcagcagctatcgctttacacatttttgaggcctactca 1200
Qy 2644 ctctagacaggaagtgggatatgccaaggctcagcttttagcaggtatttagcattttgctt 2703
Db 1201 ctctagacaggaagtgggatatgccaaggctcagcttttagcaggtatttagcattttgctt 1260
Qy 2704 ctctcatagtgtaggaagaggcggttttaaaatgctcaagtttctgattgttgacatgggg 2763
Db 1261 ctctcatagtgtaggaagaggcggttttaaaatgctcaagtttctgattgttgacatgggg 1320
Qy 2764 ctgcgaaacagatcggccagacatgattattttacagatccagatggtaccagctctcg 2823
Db 1321 ctgcgaaacagatcggccagacatgattattttacagatccagatggtaccagctctcg 1380
Qy 2824 aggtgtcttcatgattaccacagagacactcaaatcacctggaggagcagcagatcggc 2883
Db 1381 aggtgtcttcatgattaccacagagacactcaaatcacctggaggagcagcagatcggc 1440
Qy 2884 ccagcctctacgctgcccctgttctcctcaccatgttttgcctcacagttcccgcgtggga 2943

Db 1441 cccagctctacgctgccccctggttctctcaccatggttgcctcacaggtcccgctggga 1500
Qy 2944 ttcgtaccagagctctttgatgatattttcttcagggaacagaggtcattttaagtg 3003
Db 1501 ttcgtaccagagctctttgatgatattttcttcagggaacagaggtcattttaagtg 1560
Qy 3004 gctttaagctgttggaagccataagcccttgattctgcagcatgaaacccctagaacc 3063
Db 1561 gctttaagctgttggaagccataagcccttgattctgcagcatgaaacccctagaacc 1620
Qy 3064 atagttgactttataaaagcagcgtaccaccccttgcttggttacagatggaaagacc 3123
Db 1621 atagttgactttataaaagcagcgtaccaccccttgcttggttacagatggaaagacc 1680
Qy 3124 atcaatcaggtatttgaaatggacatcgctaaacagttacaagcttatgaagttgagtac 3183
Db 1681 atcaatcaggtatttgaaatggacatcgctaaacagttacaagcttatgaagttgagtac 1740
Qy 3184 cagctctcaagaagaacttatcgattctctctctcgtgacacccaaagaatggat 3243
Db 1741 cagctctcaagaagaacttatcgattctctctcgtgacacccaaagaatggat 1800
Qy 3244 aaattagaaaaaacacagcagcttacgcaaacagaaaccttgccttgaaacagttg 3303
Db 1801 aaattagaaaaaacacagcagcttacgcaaacagaaaccttgccttgaaacagttg 1860
Qy 3304 caggtgcaaatggtagatccaaagccttgaggccaccattgagaagctccttgagcagt 3363
Db 1861 caggtgcaaatggtagatccaaagccttgaggccaccattgagaagctccttgagcagt 1920
Qy 3364 gagagcaagctgaagcaggccatgcttaaccttagaactgagcgtgcccctgctgcag 3423
Db 1921 gagagcaagctgaagcaggccatgcttaaccttagaactgagcgtgcccctgctgcag 1980
Qy 3424 acggtgagagctgcgcggcgagcgcagagcccgagcggagcctgagctgagtcagc 3483
Db 1981 acggtgagagctgcgcggcgagcgcagagcccgagcggagcctgagctgagtcagc 2040
Qy 3484 cagcccgagcccgagcggcagctgacagctgctgcaggagagattgcaacaccccccac 3543
Db 2041 cagcccgagcccgagcggcagctgacagctgctgcaggagagattgcaacaccccccac 2100
Qy 3544 tgtccagggcctt 3555
Db 2101 tgtccagggcctt 2112

RESULT 8
AAS34787
ID AAS34787 standard; cDNA; 1422 BP.
XX AC AAS34787;
XX XX
DT 04-DEC-2001 (first entry)
XX XX
DE cDNA encoding novel human neoplastic disease associated polypeptide #21.
XX XX
KW Human; neoplastic disease associated polypeptide; cancer; gene therapy;
KW hyperproliferative disorder; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW neuroprotective; cytostatic; anti inflammatory; vasotropic; ss.
XX XX
OS Homo sapiens.
XX XX
PN W0200155163-A1.
XX XX
PD 02-AUG-2001.
XX XX
PF 17-JAN-2001; 2001WO-US01358.
XX XX
PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 30-JUN-2000; 2000US-0214886.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.

02-OCT-2000;	2000US-0237038	PT	arthritis
02-OCT-2000;	2000US-0237039	XX	
02-OCT-2000;	2000US-0237040	PS	Claim 4; SEQ ID No 31; 687pp; English.
13-OCT-2000;	2000US-0239935	XX	
13-OCT-2000;	2000US-0239937	CC	The present invention relates to the isolation of novel human neoplastic
20-OCT-2000;	2000US-0240960	CC	disease associated polypeptides (AAU21568-AAU21851), and cDNA and DNA
20-OCT-2000;	2000US-0241221	CC	sequences encoding for these polypeptides. The sequences of the
20-OCT-2000;	2000US-0241785	CC	invention are useful in the diagnosis, treatment, prevention and/or
20-OCT-2000;	2000US-0241786	CC	prognosis of disorders involving neoplastic disease such as
20-OCT-2000;	2000US-0241787	CC	hyperproliferative disorders (e.g. leukemia, bone cancer, bladder
20-OCT-2000;	2000US-0241808	CC	cancer, brain stem glioma, adult liver cancer, childhood cerebellar
20-OCT-2000;	2000US-0241809	CC	astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
20-OCT-2000;	2000US-0241826	CC	also be useful for treating other disorders such as neural disorders,
01-NOV-2000;	2000US-0244617	CC	immune system disorders, muscular disorders, reproductive disorders,
08-NOV-2000;	2000US-0246474	CC	gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
08-NOV-2000;	2000US-0246475	CC	and renal disorders. The polynucleotide sequences of the invention are
08-NOV-2000;	2000US-0246476	CC	also useful in gene therapy. AAS34767-AAS35050 represent cDNA sequences
08-NOV-2000;	2000US-0246477	CC	encoding for the novel human neoplastic disease associated polypeptides
08-NOV-2000;	2000US-0246478	CC	of the invention.
08-NOV-2000;	2000US-0246523	CC	Note: The sequence data for this patent did not form part of the printed
08-NOV-2000;	2000US-0246524	CC	specification, but was obtained in electronic format directly from WIPO
08-NOV-2000;	2000US-0246525	CC	at ftp.wipo.int/pub/published_pct_sequences.
08-NOV-2000;	2000US-0246526	XX	
08-NOV-2000;	2000US-0246527	SQ	Sequence 1422 BP; 399 A; 322 C; 337 G; 364 T; 0 Other;
08-NOV-2000;	2000US-0246528		
08-NOV-2000;	2000US-0246532		
08-NOV-2000;	2000US-0246609		
08-NOV-2000;	2000US-0246610		
08-NOV-2000;	2000US-0246611		
08-NOV-2000;	2000US-0246613		
17-NOV-2000;	2000US-0249207	Qy	Query Match 28.2%; Score 1000.8; DB 22; Length 1422;
17-NOV-2000;	2000US-0249208	Db	Best Local Similarity 99.3%; Pred. No. 1.5e-262;
17-NOV-2000;	2000US-0249209	Qy	Matches 1005; Conservative 0; Mismatches 7; Indels 0; Gaps 0
17-NOV-2000;	2000US-0249210	Db	
17-NOV-2000;	2000US-0249211	Qy	2544 tcttattgacctggggcgaaacctttctctacacacccatacttctgtccagattggagc 2603
17-NOV-2000;	2000US-0249212	Db	139 tattctttccagggcgaaacctttctctacacacccatacttctgtccagattggagc 198
17-NOV-2000;	2000US-0249213	Qy	2604 aggacgctatcgctttacacattttgaaggcctactctctacacaggaagtggg 2663
17-NOV-2000;	2000US-0249214	Db	199 aggacgctatcgctttacacattttgaaggcctactctctacacaggaagtggg 258
17-NOV-2000;	2000US-0249215	Qy	2664 atattgccaaggctcagctttgtagcaggcattttgtcttctctatgatgaggaaga 2723
17-NOV-2000;	2000US-0249216	Db	259 atattgccaaggctcagctttgtagcaggcattttgtcttctctatgatgaggaaga 318
17-NOV-2000;	2000US-0249217	Qy	2724 ggcgtttaaaaatgctcaagtttctgtatgtttgacatggggctcggaaacagatcgcc 2783
17-NOV-2000;	2000US-0249245	Db	319 ggcgtttaaaaatgctcaagtttctgtatgtttgacatggggctcggaaacagatcgcc 378
17-NOV-2000;	2000US-0249264	Qy	2784 agacatgattattttacagatccagatgtaccagctctcgaggtgtcttcattacc 2843
17-NOV-2000;	2000US-0249297	Db	379 agacatgattattttacagatccagatgtaccagctctcgaggtgtcttcattacc 438
17-NOV-2000;	2000US-0249299	Qy	2844 cagagacctctacaatcacctggaggagcagcagatggtggccagcgtctacgctgcc 2903
01-DEC-2000;	2000US-0250160	Db	439 cagagacctctacaatcacctggaggagcagcagatggtggccagcgtctacgctgcc 498
05-DEC-2000;	2000US-0251030	Qy	2904 ctggttctcaccatgtttgcctcacaggttcccgctgggattcgtgacagagctttga 2963
05-DEC-2000;	2000US-0256719	Db	499 ctggttctcaccatgtttgcctcacaggttcccgctgggattcgtgacagagctttga 558
06-DEC-2000;	2000US-0251479	Qy	2964 tatgattttcttcagggaacagaggttcattattaaagtggcttaagtctgttgggaag 3023
08-DEC-2000;	2000US-0251856	Db	559 tatgattttcttcagggaacagaggttcattattaaagtggcttaagtctgttgggaag 618
08-DEC-2000;	2000US-0251869	Qy	3024 ccataagcccttgattctgcagcatgaaacacctagaaacccatagttgactttataaaaag 3083
08-DEC-2000;	2000US-0251989	Db	619 ccataagcccttgattctgcagcatgaaacacctagaaacccatagttgactttataaaaag 678
08-DEC-2000;	2000US-0251990	Qy	3084 cactgtaaccacaccttgcttggtgtacagatggaaagaccatacaatcaggtatttgaat 3143
11-DEC-2000;	2000US-0254097	Db	679 cactgtaaccacaccttgcttggtgtacagatggaaagaccatacaatcaggtatttgaat 738
05-JAN-2001;	2001US-0259678	Qy	3144 ggacatcgctaaacagttacaaagcttatgaattgagttgagtcacacgtctctcaagaagaact 3203
(HUMA-) HUMAN GENOME SCI INC.		Db	739 ggacatcgctaaacagttacaaagcttatgaattgagttgagtcacacgtctctcaagaagaact 798
Rosen CA, Barash SC, Ruben SM;		XX	
WPI: 2001-465558/50.		XX	
P-PSDB; AAU21588.		XX	
Novel polypeptides and polynucleotides useful as diagnostic reagents to			

PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 XX WPI; 2001-465558/50.
 DR P-FSDB; AAU21586.
 XX
 PT Novel polypeptides and polynucleotides useful as diagnostic reagents to
 PT diagnose diseases or disorders associated with aberrant expression or
 PT activity of polypeptides, and for treating cancers, rheumatoid
 PT arthritis
 XX
 PS Claim 4; SEQ ID No 29; 687pp; English.
 XX
 CC The present invention relates to the isolation of novel human neoplastic
 CC disease associated polypeptides (AAU21568-AAU21851), and cDNA and DNA
 CC sequences encoding for these polypeptides. The sequences of the
 CC invention are useful in the diagnosis, treatment, prevention and/or
 CC prognosis of disorders involving neoplastic disease such as
 CC hyperproliferative disorders (e.g. leukemia, bone cancer, bladder
 CC cancer, brain stem glioma, adult liver cancer, childhood cerebellar
 CC astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
 CC also be useful for treating other disorders such as neural disorders,
 CC immune system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
 CC and renal disorders. The polynucleotide sequences of the invention are
 CC also useful in gene therapy. AAS34767-AAS35050 represent cDNA sequences
 CC encoding for the novel human neoplastic disease associated polypeptides
 CC of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1269 BP; 362 A; 313 C; 277 G; 312 T; 5 other;

 Query Match 26.5%; Score 943.2; DB 22; Length 1269;
 Best Local Similarity 98.7%; Pred. No. 7.7e-247;
 Matches 967; Conservative 3; Mismatches 8; Indels 2; Gaps 2;

 QY 1908 aaaggacttgaatccaagaacaaacattcttggtgattcttggtgggactctgtgaaagac 1967
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 292 aagggaacttgaatccaagaacaaacattcttggtgattcttggtgggactctgtgaaagac 351
 QY 1968 ccgagggaattcttgaggcagcagatattcttcgagtagcaccgcccgagaaagcgctg 2027
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 352 ccgagggaattcttgaggcagcagatattcttcgagtagcaccgcccgagaaagcgctg 411
 QY 2028 cgattcttcacgagatgaagattattcagagctggagagcttccccacgactccc 2087
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 412 cgattcttcacgagatgaagattattcagagctggagagcttccccacgactccc 471

QY 2088 tttagaacacagtttggatgaagatgggcccctttggccccccacacagaggaagaaagac 2147
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 472 tttagaacacagtttggatgaagatgggcccctttggccccccacacagaggaagaaagac 531
 QY 2148 atctcgtgagctccgagagctgtggcaaaagctatttctcaacagatactgctcttag 2207
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 532 atctcgtgagctccgagagctgtggcaaaagctatttctcaacagatactgctcttag 591
 QY 2208 aatggagaagaaatcagaagctccaaagctctgaaatgatttgcgtgaacaagcgctc 2267
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 592 aatggagaagaaatcagaagctccaaagctctgaaatgatttgcgtgaacaagcgctc 651
 QY 2268 gaagctcattatgaagaaattactccctgtcttaaaagaagtaactacagtggtggaata 2327
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 652 gaagctcattatgaagaaattactccctgtcttaaaagaagtaactacagtggtggaata 711
 QY 2328 gatccttagcactccaggaagatcaaaaataaagtttgacatggaaaatacgcactcgc 2387
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 712 gatccttagcactccaggaagatcaaaaataaagtttgacatggaaaatacgcactcgc 771
 QY 2388 tgttgggcaaggtgtgccacgtcattccacggaggtgaaatctggaaaatttctagctgagca 2447
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 772 tgttgggcaaggtgtgccacgtcattccacggaggtgaaatctggaaaatttctagctgagca 831
 QY 2448 attccaccttaaacaccagtttccacagcaaacagcagccaaagagtggtgccatacaaga 2507
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 832 attccaccttaaacaccagtttccacagcaaacagcagccaaagagtggtgccatacaaga 891
 QY 2508 actctaaagcagctgacttcccagcagcatgcgattcttattgaccttggcggaacctt 2567
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 892 actctaaagcagctgacttcccagcagcatgcgattcttattgaccttggcggaacctt 951
 QY 2568 tcttacacaccatacttctctgcccagcttggagcagcagctatgcgttacaacat 2627
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 952 tcttacacaccatacttctctgcccagcttggagcagcagctatgcgttacaacat 1011
 QY 2628 tttaaggccttactcactcttagaccaggaagtggtggtatgcccaggtcttcagctttgt 2687
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1012 tttaaggccttactcactcttagaccaggaagtggtggtatgcccaggtcttcagctttgt 1071
 QY 2688 agcaggccttcttcttcttctatagtgaggaagagcggtttaaatactcaagttct 2747
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1072 agcaggccttcttcttcttctatagtgaggaagagcggtttaaatactcaagttct 1131
 QY 2748 gatgtttacatggggctgcggaaacagatgcggccagacatgattattttacagatcca 2807
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1132 gatgtttacatggggctgcggaaacagatgcggccagac-tgattattttacagatcca 1190
 QY 2808 gatgtaccagctctcgaggttgccttcattgattaccacagagacctctacaatcacctgga 2867
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1191 gatgtwccagctctcgaggttgccttcattgattaccacagana-ctctaccatcacctgga 1249
 QY 2868 ggagcagcagatcgggcccca 2887
 DB ||||||||||||||||
 DB 1250 agaacacgaatcngggcca 1269

 RESULT 11
 AAH04450
 ID AAH04450 standard; cDNA; 883 BP.
 XX
 AC AAH04450;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (5'-primer) SEQ ID NO:1285.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX

PD 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

PF 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

PI WPI; 2001-318749/34.

XX

XX Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

XX

PS Claim 1; SEQ ID 1285; 2537pp + CD ROM; English.

XX

CC The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX

XX Sequence 883 BP; 245 A; 221 C; 219 G; 193 T; 5 other;

SQ

Query Match 22.1%; Score 785; DB 22; Length 883;

Best Local Similarity 96.7%; Pred. No. 9.9e-204;

Matches 843; Conservative 0; Mismatches 23; Indels 6; Gaps 4;

Qy 1444 agtaattaccacccagtcgcatctgatttagctagatattgtgaaacaaagcaag 1503

Db 1 agtaattaccacccagtcgcatctgatttagctagatattgtgaaacaaagcaag 60

Qy 1504 agatctttaacagagcttttagaagattttgtcccggttaataaagccagagcctg 1563

Db 61 agatctttaacagagcttttagaagattttgtcccggttaataaagccagagcctg 120

Qy 1564 caggaacacccatcagctggtatctgtagatctccctgtctctagatcattaaataaccc 1623

Db 121 caggaacacccatcagctggtatctgtagatctccctgtctctagatcattaaataaccc 180

Qy 1624 agcaagagccatctgtgtgaaagagggccttgcctctctgagagctctttaaag 1683

Db 181 agcaagagccatctgtgtgaaagagggccttgcctctctgagagctctttaaag 240

Qy 1684 ctctcggtctctcgagagaccctgtccagtgactcggagagtgactctccagagagccca 1743

Db 241 ctctcggtctctcgagagaccctgtccagtgactcggagagtgactctccagagagccca 300

Qy 1744 gctccgctgtgccccagcagcagccttcagaggcgagagcaaacacccctgagctactcccc 1803

Db 301 gctccgctgtgccccagcagcagccttcagaggcgagagcaaacacccctgagctactcccc 360

Qy 1804 atcgaatgcaggaacacctccacaacctgcccgggggtcccccgggggttttcgcaaaagaaa 1863

Db 361 atcgaatgcaggaacacctccacaacctgcccgggggtcccccgggggttttcgcaaaagaaa 420

Qy 1864 cttatgaggtatcactcagtgagcagacagagcctcatgaacgaagagactttgaaatcc 1923

Db 421 cttatgaggtatcactcagtgagcagacagagcctcatgaacgaagagactttgaaatcc 480

Qy 1924 aaagcaaacatctgtgtgactctgtgagactcctctgtaagaccccgagagcattctcg 1983

Db 481 aaagcaaacatctgtgtgactctgtgagactcctctgtaagaccccgagagcattctctgg 540

Qy 1984 aggcagcagatatctcccgagtagccaccccgcaagaaggcgtgcgattctctccagcaga 2043

Db 541 aggcagcagatatctcccgagtagccaccccgcaagaaggcgtgcgattctctccagcaga 600

Qy 2044 tatgaagattattcagagctgggagagcttccccccacgactctctttagaaccagttgt 2103

Db 601 tatgaagattattcagagctgggagagcttccccccacgactctctttagaaccagttgt 659

Qy 2104 gaagatggccctttggccccccaccagagagaaagagacatctcgtgagctccga 2163

Db 660 gaagatggccctttggccccccaccagagagaaagagacatctcgtgagctccga 719

Qy 2164 gagc-tgtggcaaaagctattcttcaacagatactgctgcttagaattgagagagaaa 2222

Db 720 agcttgtggcaaaagctattcttcaacagatactgctgcttagaattgagagagaaa 779

Qy 2223 taagaagctccaaagcc---tctgaaatgatttgcgtgaacaaagcgc-tgaagctcgatt 2278

Db 780 taagaagctccaaagcctttgaaatgatttgcgtgaacaaagcgccttgaactcgatt 839

Qy 2279 atgaagaataattactcctcctcttaagaagagta 2310

Db 840 ntgaagaataattactcctcctttaaagaagta 871

RESULT 12

AAK92006

ID AAK92006 standard; cDNA; 849 BP.

XX AAK92006;

XX 06-NOV-2001 (first entry)

XX Human cDNA 5'-end sequence, SEQ ID NO: 466.

DE Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

KW Homo sapiens.

OS EP1130094-A2.

PN 05-SEP-2001.

XX 07-JUL-2000; 2000EP-0114089.

XX 08-JUL-1999; 99JP-0194486.

PR 11-JAN-2000; 2000JP-0118774.

PR 02-MAY-2000; 2000JP-0183765.

XX (HELI-) HELIX RES INST.

PA Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX WPI; 2001-524255/58.

PT 830 Primers useful for synthesizing full length cDNA clones and their
XX use in genetic manipulation -
PS Claim 2; SEQ ID NO 466; 1380pp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones, 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is the nucleotide
CC sequence of the 5'-end of a cDNA provided in the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 849 BP; 224 A; 195 C; 196 G; 231 T; 3 other;

Query Match 16.2%; Score 574.2; DB 22; Length 849;
Best Local Similarity 98.2%; Pred. No. 3.5e-146;
Matches 642; Conservative 0; Mismatches 6; Indels 6; Gaps 6;
Qy 2555 ttggcggaacatttctctacacacccatactctctgcccagcttggagcagacagctat 2614
Db 198 ttggcggaacatttctctacacacccatactctctgcccagcttggagcagacagctat 257
Qy 2615 cgtcttaacaacatttgaagcctactcactctctagaccaggaagtgggatattgccaag 2674
Db 258 cgtcttaacaacatttgaagcctactcactctctagaccaggaagtgggatattgccaag 317
Qy 2675 gtctcagctttgtagcagcatttggcttctctatagtgaggaagagcggttttaaaa 2734
Db 318 gtctcagctttgtagcagcatttggcttctctatagtgaggaagagcggttttaaaa 377
Qy 2735 tgcctcagctttgtagcagcatttggcttctctatagtgaggaagacatgatta 2794
Db 378 tgcctcagctttgtagcagcatttggcttctctatagtgaggaagacatgatta 437
Qy 2795 tttacagatccagatgaccagctctcgaggttgcttctctatagtgaggaagacatgatta 2854
Db 438 tttacagatccagatgaccagctctcgaggttgcttctctatagtgaggaagacatgatta 497
Qy 2855 acaatcaactggaggagacagagatcgcccgccagcctctacgtcgccctggttctctca 2914
Db 498 acaatcaactggaggagacagagatcgcccgccagcctctacgtcgccctggttctctca 557
Qy 2915 ccattgttgcctcagcttcccgctgggttcgttagcagagcttcttgatgatttttc 2974
Db 558 ccattgttgcctcagcttcccgctgggttcgttagcagagcttcttgatgatttttc 617
Qy 2975 ttcagggaacagaggtcatattttaaagtggctttaaagctgttgggaagccataagccct 3034
Db 618 ttcagggaacagaggtcatattttaaagtggctttaaagctgttgggaagccataagccct 677
Qy 3035 tgattctgcagcatgaanaacctagaacacatagttgactttataaaaagcagcgtaccca 3094
Db 678 tgattctgcagcatgaanaacctagaacacatagttgactttataaaaagcagcgt-ccca 736
Qy 3095 accttggcttggtacagatgaaagaccatcaatcaatcaggtta-tttgaaatggacatc-gc 3152
Db 737 accttggcttggt-acagatgaaagaccatcaatcaatcaggttaatttgaatggacatcgg 795
Qy 3153 taaacagttacaagcttatg-aagttaagttacca-cgtcccttcaagaagaactt 3204
Db 796 taaacagttacaagcttatgaaagttaagttaccacgttccctttaaagaagaactt 849

RESULT 13

AAK93666

ID AAK93666 standard; cDNA; 849 BP.

XX AAK93666;
XX 06-NOV-2001 (first entry)
XX Human cDNA clone representative sequence, SEQ ID NO: 2126.
DE
XX
DE Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
KW
XX
XX Homo sapiens.
OS
XX
XX EPI130094-A2.
PN
XX
XX 05-SEP-2001.
PD
XX
XX 07-JUL-2000; 2000EP-0114089.
PF
XX
XX 08-JUL-1999; 99JP-0194486.
PR
XX
XX 11-JAN-2000; 2000JP-0118774.
PR
XX
XX 02-MAY-2000; 2000JP-0183765.
PR
XX
XX (HELI-) HELIX RES INST.
PA
XX
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
PI
XX
XX WPI; 2001-524255/58.
DR
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
PT
XX
XX Example 11; SEQ ID NO 2126; 1380pp + sequence listing; English.
PS
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence was used as the
CC representative sequence from a human clone which was used in
CC homology searches to identify the clone.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 849 BP; 224 A; 195 C; 196 G; 231 T; 3 other;

Query Match 16.2%; Score 574.2; DB 22; Length 849;
Best Local Similarity 98.2%; Pred. No. 3.5e-146;
Matches 642; Conservative 0; Mismatches 6; Indels 6; Gaps 6;

Qy 2555 ttggcggaacatttctctacacacccatactctctgcccagcttggagcagacagctat 2614
Db 198 ttggcggaacatttctctacacacccatactctctgcccagcttggagcagacagctat 257
Qy 2615 cgtcttaacaacatttgaagcctactcactctctagaccaggaagtgggatattgccaag 2674
Db 258 cgtcttaacaacatttgaagcctactcactctctagaccaggaagtgggatattgccaag 317
Qy 2675 gtctcagctttgtagcagcatttggcttctctatagtgaggaagagcggttttaaaa 2734
Db 318 gtctcagctttgtagcagcatttggcttctctatagtgaggaagagcggttttaaaa 377
Qy 2735 tgcctcagctttgtagcagcatttggcttctctatagtgaggaagacatgatta 2794
Db 378 tgcctcagctttgtagcagcatttggcttctctatagtgaggaagacatgatta 437
Qy 2795 tttacagatccagatgaccagctctcgaggttgcttctctatagtgaggaagacatgatta 2854
Db 438 tttacagatccagatgaccagctctcgaggttgcttctctatagtgaggaagacatgatta 497

QY 181 accaagcaagtcgctgttgcgtttcaccctctgactgagatggaacccagcagg 240
DB 13013 accaagcaagtcgctgttgcgtttcaccctctgactgagatggaacccagcagg 13072
QY 241 aagaatcaacagtgagtcctcctgactatctcagcatcttggatgcaagcctcagcgt 300
DB 13073 aagaatcaacagtgagtcctcctgactatctcagcatcttggatgcaagcctcagcgt 13132
QY 301 gttccaaactgattccaaacagtcagtcacccaagtactttgctgtctgattaaggaa 360
DB 13133 gttccaaactgattccaaacagtcagtcacccaagtactttgctgtctgattaaggaa 13192
QY 361 gacgtgtccaccgacagatgctgtatgtgttcaaacgcatgacaaacaaagt 420
DB 13193 gacgtgtccaccgacagatgctgtatgtgttcaaacgcatgacaaacaaagt 13252
QY 421 cctgagatcattcagtcctcctcgtcagcgggg 453
DB 13253 agtgagatggagatccaaagactaagtggtgg 13285

RESULT 15
ID AA215229 standard; cDNA; 696 BP.
XX AA215229;
AC AA215229;
XX 12-OCT-1999 (first entry)
XX Human gene expression product cDNA sequence SEQ ID NO:2698.
XX Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX Homo sapiens.
XX WO9938972-A2.
XX 05-AUG-1999.
XX 28-JAN-1999; 99WO-US01619.
XX 03-APR-1998; 98US-0080666.
XX 28-JAN-1998; 98US-0072910.
XX 24-FEB-1998; 98US-0075954.
XX 31-MAR-1998; 98US-0080114.
XX 03-APR-1998; 98US-0080515.
XX (CHIR) CHIRON CORP.
XX (HYSE-) HYSEQ INC.

PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX WPI; 1999-494092/41.
XX Novel human genes and their expression products which are
PT differentially expressed in different cell types
XX
XX Claim 1; Page 1311; 2479pp; English.

XX The present invention describes a library of human polynucleotides
CC comprising the sequences given in AA215232 to AA21779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AA215232 to AA21779. The
XX polynucleotides can be used as a source of primers and probes, which can

CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
SQ Sequence 696 BP; 189 A; 157 C; 178 G; 142 T; 30 other;

Query Match 10.2%; Score 362; DB 20; Length 696;
Best Local Similarity 100.0%; Pred. No. 2.8e-88;
Matches 362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3194 aagaagaacttatcctctcctcctcagtgacacacaaagaatggataaattagaga 3253
DB 59 aagaagaacttatcctcctcctcctcagtgacacacaaagaatggataaattagaga 118
QY 3254 aaaccaacagcagcttacgcaaaacagacaccttgacctccttgacagcttgaggca 3313
DB 119 aaaccaacagcagcttacgcaaaacagacaccttgacctccttgacagcttgaggca 178
QY 3314 atggtagatccaaagccttgagccaccattgagaagctcctgagcagtgagcaagc 3373
DB 179 atggtagatccaaagccttgagccaccattgagaagctcctgagcagtgagcaagc 238
QY 3374 tgaagcagggccatgcttacccttagaactggagcgtcgccctgctgacagcgtggagg 3433
DB 239 tgaagcagggccatgcttacccttagaactggagcgtcgccctgctgacagcgtggagg 298
QY 3434 agctcggcggcggagcagcagccagcagcggggagcctgagtcagcagcccgagc 3493
DB 299 agctcggcggcggagcagcagccagcagcggggagcctgagtcagcagcccgagc 358
QY 3494 ccacggcgactgacagctctgacagagagattgcaacacccatccacactgtccaggcc 3553
DB 359 ccacggcgactgacagctctgacagagagattgcaacacccatccacactgtccaggcc 418
QY 3554 tt 3555
DB 419 tt 420

Search completed: September 9, 2002, 19:20:22
Job time: 17600 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2002, 19:16:17 ; Search time 8477.27 Seconds
(without alignments)
8775.683 Million cell updates/sec

Title: US-09-762-311-4_COPY_176_3730

Perfect score: 3555

Sequence: 1 atgggaacaataacattcac.....tccacactgtccaggcctt 3555

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hcg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pi.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pi.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	2751	77.4	3023	9	BC014529
2	2510.4	70.6	4039	6	186429
3	2510.4	70.6	4039	10	MM033005
4	2340	65.8	2576	9	AB029031
5	2112	59.4	2362	9	AK027355
6	1359.2	38.2	1995	4	BT017923
7	1179.6	33.2	3051	10	BC004675
8	598.4	28.1	1788	9	AK057182
9	682.2	19.2	5922	9	AB011175
10	465.8	13.1	189811	30	AC009595
11	465.8	13.1	195108	9	AC021106
12	465.8	13.1	208318	2	AC108933
13	424.2	11.9	110000	2	AL390202_04
14	304.6	8.6	3214	3	DM050542
15	304.6	8.6	4746	3	DME17919
16	284.4	8.0	189811	30	AC009595
17	284.4	8.0	208318	2	AC108933
18	252.4	7.1	163928	2	AC021335
19	252.4	7.1	180918	9	AC098680
20	245.6	6.9	17019	2	AC019886
21	245.6	6.9	161117	3	AC008311
22	245.6	6.9	168448	3	AC009346
23	245.6	6.9	297832	3	AE003602
24	180	5.1	163928	2	AC021335
25	141.8	4.0	3761	9	BC014111
26	141	4.0	221717	2	AC074357
27	139.2	3.9	158310	9	AL162571
28	139.2	3.9	183577	2	AC011164
29	132.8	3.7	3691	6	AX078376
30	132.8	3.7	4560	9	HS011679
31	123.8	3.5	154	6	AX261661
32	119.4	3.4	14119	2	AC013898
33	119.4	3.4	184413	3	AC023718
34	119.4	3.4	303648	3	AE003486
35	114.6	3.2	2013	9	AF042345
36	114.6	3.2	7383	9	AF008915
37	101.8	2.9	30027	3	CELT24D11
38	95.6	2.8	20553	3	CBRG26M15
39	95.6	2.7	73803	8	NCB11B22
40	84.2	2.5	5870	10	MM053586
41	84.2	2.4	441	11	G47994
42	80	2.3	27102	2	U41540
43	79.6	2.2	110032	2	AC014916
44	79.6	2.2	159660	3	AC010040
45	79.6	2.2	282267	3	AE003555

ALIGNMENTS

RESULT 1

BC014529

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

BC014529 3023 bp mRNA linear PRI 19-DEC-2001
Homo sapiens, clone IMAGE:3843156, mRNA.

BC014529

BC014529.1 GI:17939551

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 3023)

Strausberg,R.

Direct Submission

Submitted (24-SEP-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Institute for Systems Biology
 http://www.systemsbio.org
 Contact: amadanes@systemsbiology.org
 Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Series: IRAL Plate: 14 Row: c Column: 16
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis.

FEATURES

source

Location/Qualifiers

1..3023
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3843156"
 /tissue_type="placenta, choriocarcinoma"
 /clone_lib="NIH_MGC_21"
 /lab_host="pH10B-R"
 /note="Vector: pORB7"
 BASE COUNT 883 a 712 c 740 g 688 t

Query Match 77.4%; Score 2751; DB 9; Length 3023;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 805 attgagaacacacctattagcgacacacatttgcagccacacagatatcgaggaat 864
 DB 10 ATTGAGAACCCCTCATTAGCGACACATATTGTGCGCCCGACACATATCGAGGAAAT 69
 QY 865 cgaactatgctcttcacgattggccagtcgtgaagtttaacctatcactgctgacaccaa 924
 DB 70 CGAACTATGCTCTTCAGATTGGCCAGTCTGAAGTTTACCTCATCTAGCTGTCGACCAAA 129
 QY 925 aaaaatgacttagagaaaatttaaggagatatctttgctctcagggcactcagacac 984
 DB 130 AAAATAGCATGTGAGAAAATTTTAAAGGAGATATCTTTTGTCTCAGGCGATCAGACAC 189
 QY 985 gtggaccactttgggtttatctgcgggagcttcgggaggtggcggttcaatttgc 1044
 DB 190 GTGGACCACTTTGGGTTTATCTCTGGGAGCTCTTCGGAGGTGGCGCTTTCATTTTGTC 249
 QY 1045 tgttacggttttcagtgacaaaataggctctggttgatgaaattatgatgacctgaaa 1104
 DB 250 TGTTAGCTGTTTCAGTGCACAAATGAGGCTCTGGTTGATGAAATTTATGATGACCCCTGAAA 309
 QY 1105 caggcttcacggtggccagtcgacgacagctaaagcgccagccagctgtgtgag 1164
 DB 310 CAGGCTTTCAGGTGGCCGCGAGTGCAGACAGACAGCTAAGCGCCGACCCAGCTGTGTGAG 369
 QY 1165 ggtgccccctgcaagcctgcacagctctgtgagagatagagggaaatcttcc 1224
 DB 370 GGCTGCCCTCTGCAAGCCCTGCACAGCTCTGTGAGAGGATAGAGGAATGATTTCTTCC 429
 QY 1225 aaacaaaactagaactgcaaaagcaccctgacgacatttaaccaatcagagcaggcgact 1284
 DB 430 AAAACAAAATAGAACTGCAAAAGCACCTGACGACATTTAACCAATCAGGAGCAGGCGACT 489
 QY 1285 atttttgaagaggttcagaaattgagccgagaaatgagcagcagagagaaatgatt 1344
 DB 490 ATTTTGAAGAGGTTTCAAAAATTTGACCCGAGAAATGAGCAGCGAGAGATGAATTTGATT 549
 QY 1345 attctttctgagattttatgtgaagaaacacagaaacacatccatattggggag 1404
 DB 550 ATTTCTTTCTGAGATGTTTATATGAGAGAAACACAGAAAGACACATATATTTGGGAG 609
 QY 1405 atgaagcagacatcgagatgagcagcagagaaatattggagtgaaataccaccagtgcc 1464
 DB 610 ATGAAGCAGACATCGGAGATGGCAGCAGAGATAATTGGGAAGTGAATTAACCCAGCTGCC 669

QY 1465 actcatttagcgtagatatgctgaaaaaacaagcaaaagatctttaacagagtcttta 1524
 DB 670 ACTCGATTTAGCTAGATATGCTGAAAAACAAGCAAAAGAGATCTTTACAGAGTCTTTA 729
 QY 1525 gaaagtattttctcccggggtaataaaagccagaggcctgcaggaacacactccatcagtgtg 1584
 DB 730 GAAAGTATTTTCTCCCGGGTAAATAAGCCAGAGGCTGCAGGAACACTCCATCAGTGTG 789
 QY 1585 gatctgatatgctccctctctagtaacattaaacacacagcaaaagcagcactctgtgtg 1644
 DB 790 GATCTGATATAGTCCCTGTCTAGTACATTAAGTAACACAGCAAGAGGACATCTGTGTGT 849
 QY 1645 gaaagaggccttgcccatctctgagagctctctttaaagctctctcggtcctcgaggagac 1704
 DB 850 GAAAGGAGGCTTGGCCATCTCTGAGAGCTCCTTTAAGCTCCTCGGCTCCTCGGAGGAC 909
 QY 1705 ctgtccagtgactcggagagtgatctctccagagagccagctcctgtctgcgccccagag 1764
 DB 910 CTGTCCAGTGACTCGGAGAGTCTCTCCGAGAAGAGCCAGCTCCGCTGTCCGCCAGCAG 969
 QY 1765 gcttcaggagcgagcaaaacacctgagtcactctcccatcgaaatgccaggaaacctcca 1824
 DB 970 GCCTTCAGAGGCGAGCAAAACACCTCTGATCTCTCCCATCGAATGCCAGGAACCTCCA 1029
 QY 1825 caactgccccggggtccccgggggttcgcaaaaggaaacttatgaggtatcactcagtgt 1884
 DB 1030 CAACCTGCCCGGGGTCCCGGGGTTTCGCAAAAGAAACTTATGAGGTATCACTCAGTGT 1089
 QY 1885 agcacagagacgcctcatgaacgaaggactttgaatccaaagcaaacactcttggatgt 1944
 DB 1090 AGCACAGAGACGCTCATGAACGAAGAGACTTTTGAATCCAAAGCAAACTCTTGTGTGT 1149
 QY 1945 tctgtgggaactcctgtgaagaccggagagcttccctgcagcagcagatattccctccga 2004
 DB 1150 TGTGTGGGACTCTCTGGAAGACCCGGAGCATTTCTGGAGGCGCAGATATTTCCTCCGA 1209
 QY 2005 gtggccaccgcgcagagcgctgcgattcttccagcagatatgaagattattcagagctg 2064
 DB 1210 GTAGCCACCCCGCAGAAAGCGTGCATTTCTCCAGCAGATATCAAGATTATTTCAGAGCTG 1269
 QY 2065 ggagccttccccacgatatctcttttagaacacagttgtgaaagatgggccccttggcccc 2124
 DB 1270 GGAGAGCTTCTCCCGACGATCTCTTTAGAACCACTTGTGGAAGATGGGCCCTTTGGCCCC 1329
 QY 2125 ccacagagaaaaaagagacatctcgtgagctccgagagctgtggcaaaaggctatt 2184
 DB 1330 CCACCAGAGAAAAGAAAAGGACATCTCTGAGCTCCGAGAGCTGTGGCAAAAGGCTATT 1389
 QY 2185 cttcaacagatactgctctagaatggagaaggaatacagaagctccaagcctctgaa 2244
 DB 1390 CTTCAACAGATACTGCTGTCTTAGAATGGAGAGGAAATCAGAAGCTCCAAGGCTCTGAA 1449
 QY 2245 aatgatttctgaacagcgctgaagctcgattatgaagaaattactccctctcttaa 2304
 DB 1450 AATGATTCTGTGAACAGCGCTTGAAGCTCGATTATGAAGAAATTTACTCCCTCTCTTAA 1509
 QY 2305 gaagtaactacagtggtggaaaaagatgcttagcactccaggaagatcaaaaaatgaagttt 2364
 DB 1510 GAAGTAACATACATGTGGGAAAAGACATGCTTAGCACTCCAGGAAGATCAAAAATTAAGTTT 1569
 QY 2365 gacatggaaaaaatgcactcggctgtgtggcaagggtgtgccagctcatcaccagagtga 2424
 DB 1570 GACATGAAAAAATGCACTCGGCTGTGGCAAGGTGTGCACGTCTATCACCAGGTGAA 1629
 QY 2425 atctgaaatttctagctgagcaattccaccttaaacaccagtttcccagcaaacagcag 2484
 DB 1630 ATCTGAAATTTTCTAGCTGAGCAATTTCCACTTTAAACACCAAGTTTCCCAACAAAGCAG 1689
 QY 2485 ccaagagatgtccatcacaaaagaactctaaagcagctacttcccagcagcagtgagatt 2544
 DB 1690 CCAAGGATGTGCCATACAAAAGAACTCTTAAAGCACTGACTTCCCGCAGCAGCATGCGATT 1749

QY 778 ttcttcagctcttcgagagcgacattgagaacacacctcattagcgacacacatt 837
Db 1054 TT-----TAGCTCCCTTTGACAAATGACATAGAGAACCCCTCATCGTGGGCACAATGTG 1107
QY 838 gtgagccacacagatcagagaaatacgaactatctcttcacgattggccagctctgaa 897
Db 1108 GTTCAGCCACACACATGAGAGAACCCGAACATATCTGTTACAGATTGGCCCATCTGAA 1167
QY 898 gtttaacctcatcagtcctgacacccaaaaaataagcatttgagaaaaattttaaggagata 957
Db 1168 GTTTACCTCATCATAGTCCCTGACACACAAAGATGTCACTGGAGAAAAATTTTAAGGAGATA 1227
QY 958 tctctttgctctcagggccatcagacacgtggaccactttgggttttatctgtcggagctc 1017
Db 1228 TCTCTTTGTCTCAGGGCCATCAGACATGTGGACCACTTTGGATTATCTGTCCGAGAGTGC 1287
QY 1018 tc-----cggaggtggcgcttctcattttgtctgttacgtgtttcagtgcaaaatgag 1071
Db 1288 TCGGGTGGCGGCACTGGCGCTTTTCATTTTGTCTGTACGTGTCCAGTGCAAAATGAA 1347
QY 1072 gctcgtgttgatgaataattatgatgacctgaaacagcgcttcacgtggcgagtgcaag 1131
Db 1348 GCTCTGGTTGACGAGATCATGATGACTCTGAAGCAGGCTTTACGCTAGTGCCTGGTGCAG 1407
QY 1132 cagacagctaaaggcgccagccagctgtgtgagggctgccccctgcgaagcctgcacaag 1191
Db 1408 CAGACCGCTTAAGCCACACGCCACGCTCTGTGAGGGCTGCCCTTGCAGGCCCTGCACAG 1467
QY 1192 ctctgtgagaggatagaggaatgaattcttccaaacaaaactagaaactgcaaaagcac 1251
Db 1468 CTCTCGMAAGGATAGAGGAATGAATTCATCTAAAACCAAAATTAGAATCCAGAAGCAC 1527
QY 1252 ctgacagacattaaccaatcagggcgaggcgactatttttgaagaggttcagaaattgaga 1311
Db 1528 TTGACCACACTGACCAATCAGGAGCAGGCCACCATATTCGAGGAGGTTCCAGAAATTTGAG 1587
QY 1312 ccgagaaatgagcagcagagaaatgaattgatttcttcttgagatgtttatatgaa 1371
Db 1588 CCAAGAAACGAGCAGCGAGAGATGAATTAATTTCTTCTGTGAGGTGCTATATGAA 1647
QY 1372 gagaacagaaagaaacacatccatatttggggagatgaagcagacatcgcagatggcagca 1431
Db 1648 GAGAAGCAAAAGACACAGCCACACTGGGGCGCCAAAGCAGACACATACAGGTGGCAGCA 1707
QY 1432 gagaaatttggaagtgaattaccacccagtgccactcgatttaggtagatatgtcgaaa 1491
Db 1708 GAGAAATTTGGGAGTGACCTGCCACCCAGTGCTAGCGGTTCAGGTTAGATTGCGTGAAG 1767
QY 1492 acaaaagcaaaagagatctttaaagagctctttagaaagtattttgtcccggggtaataaa 1551
Db 1768 AACAGACAAAGAGGTCTTTAAGAGAGTCCCTAGAGAGCATTTCTGTCCCGGGTAATAA 1827
QY 1552 gccagagggctgcaggaacactccaatgltgtgattcttgatagctccctgtcttagtaca 1611
Db 1828 GCCAGAGGGCTGCAGGACCAATTCGCCAGTGTGGATCTGGACACTCCACATTTCTAGTACT 1887
QY 1612 ttaagttaacacagaaagacatctgtgtgaaagaggccttgccactctctgag 1671
Db 1888 CTAAGTAACACCAAGAGAGCTGTCCATGGGTGACAAGAGGCCCTTCCCGGCTCTGTAG 1947
QY 1672 agctctcttaagctctcggctcctcggaggaacctgtccagtgactcggagatcatctc 1731
Db 1948 ACCTCCTCAAGCTCTTGGCTCTCAGATGACCTGTCCAGTGACTCAGAGGGCCACAT 2007
QY 1732 ccagaagagccagctccgctgtgtcccccagcaggcccttcaggagggcyagaaacacccctg 1791
Db 2008 GCAGAAGAGTCTGCCCTTGTTCACCCACAGCAGCGGCTTCAGAAGGAGAGACCAACCCCTG 2067
QY 1792 agtcacttcccacatgaatgcaggaacctccaacacctgcggggggtccccgggggtt 1851
Db 2068 AGTCAATTTCCAGTAGTGCCCTGGCCCTCCAGAACCTTCCCCAGAGGCTCTCCAGGGGTC 2127
QY 1852 tcgcaaaaggaaacttatgaggtatcactcagtgagcacagagagcgctctcatgaacgaaag 1911

Db 2128 TCTCAAAAGGAAACTCATGCGGTACCACTCGTGAGCACAGAGAGCGCTCATGAACCAAG 2187
QY 1912 gactttgaatccaaagcaaacccatcttgggtgattcttgggtgggactcctgtgaaagaccccg 1971
Db 2188 GACTTTGAATCCAAAGCAAAACCACTGGGTGGTGACACAGATGGGACCCCGCTGAAGACCCG 2247
QY 1972 aggcattcttgagcgacagatatctctccagtagtcacaccccgagagggcgtgcgat 2031
Db 2248 CGGCACTCTGGAGACAGCAGATATTTCTTCGANGTGGCCACTCCACAGAAAGGCTGTGTGAC 2307
QY 2032 tcttcagcagatatgaagattattcagagctgggagagcttccccacagatctctctta 2091
Db 2308 TCCCGGAGCAGATATGAAGATTATCCGAGCTGGGAGAGCTCCCTCCACGCTCCCCCTTA 2367
QY 2092 gaaccagtttgaagatggggcccttggccccccacacagagagaaaaagagacatct 2151
Db 2368 GAACCGGTGTGTAGGACCGGCCATTTGG-CCAGTACAGGAAGAAAGAGAGAGAGCTCA 2426
QY 2152 cgtgagctccgagagctgtggcaaaagctattcttcaacagatactg-ctgcttagaat 2210
Db 2427 CCGGAGCTTCGAGAGCTGTGGAAAAGGCCATCTTCGACGAGATCTCTGCTCCTCAGGAT 2486
QY 2211 ggagaggaataatcagaagctccaagcctctgaaaaatgatttgcgtgaacaagcgctgaa 2270
Db 2487 GGAGAAAGGAGAAATCAGAAGCTACAAAGCTCTGAAAACGATTTGCTGAACAAAGCCTCAA 2546
QY 2271 gctcgattatgaagaaattactccctgttcttaaaagtaactacagtggtgggaaagat 2330
Db 2547 GCTTGACTATGAAGAAATCCTCCGCTCTTAAAGAAAGTCACTACAGTGTGGGAAAGAT 2606
QY 2331 gcttagcactccaggaagatacaaaatgaatttgacatggaaaaatgcactcggctgt 2390
Db 2607 GCTTAGCACTCCAGGAAGATCCAAAATTAAGTTTGACATGGAAAAGTGCACACTGCT 2666
QY 2391 tgggcaaggtgtgccacgtcatcccgaggtgaaatcttgaaatttctagctgagcaatt 2450
Db 2667 TGGGCAAGGTGTGCCACGCTATCACCGAGGTGAGATCTGAAAATTTCTAGCTGAGCAGTT 2726
QY 2451 ccaccttaaacaccagtttcccagcaaacagcagcaaaagatgtgccaatacaaaact 2510
Db 2727 CCACCTTAAACACCCATTTCTTAGTAAACAGCAGCCAAAGAGGCTGCCCTTACAAAGACT 2786
QY 2511 cttaaagcagctgaactcccagcagcatcgattcttattgaccttggcggaaccttcc 2570
Db 2787 CTTGAAGAGAGCTGACCTCGCAGCAGCACGCCATTCATCGACCTCGGGGGAACCTTTC 2846
QY 2571 tacacaccatacttctctgccagcttggagcagcagcagctatcgctttacaacatttt 2830
Db 2847 AACACATCCATACTTCTGTGCCAGCTTGGAGCAGGTCACTGTCTACTTTTACAACTTCT 2906
QY 2631 gaaggcactcacttcttagacaggaagtggaattgccaaagctcagctcttagc 2690
Db 2907 GAAGGCTTACTCGCTTCTGGACAGGAGGTGGATGACTGCCAAGGTCTCAGCTTTGTGGC 2966
QY 2691 aggcatttggcttctctcatatgagtgaggaagggttttaaaatgcctcaagtcttgat 2750
Db 2967 AGGCATTTTGTCTTTCATGAGTGAGGAGAGGGCTTCAAGATGCTCAAGTTCCTGAT 3026
QY 2751 gttgacatggggctgcgaaacagtatcgccagacatgattattttacagatccagat 2810
Db 3027 GTTTGACATGGGCTCGGAAAACAGTATCGGGCAGACATGATTATTTTCAGATCCAGAT 3086
QY 2811 gtaccagctctcaggttgccttcattacacagagacacctcaaatcacctggagga 2870
Db 3087 GTACCAGCTGTACGCTCTCTCCACGATTACCACCGAGACCTCTACACACCTGGAAGA 3146
QY 2871 gacagagatcgcccgagcctctacgtgccccctggttctcctacacatgttggctctca 2930
Db 3147 GCACGAGACTGGCCCCCTACGTACGCGCTCCCTGGTTCTCACCGCTGTTCGCCCTCACA 3206
QY 2931 gttcccgctgggattcgtagccagagctcttgatagattttcttcagggaaacaggt 2990

Db 892 TTCAACCATGTGAGCTGTGGTCCGACAGACGGACTGGGAAGCCGCCACCGGGGAGCCATCA 951
Qy 658 GCGCCACAGGAGCAGGAGCCTGTGCGCAGGCCATGCGCAAGTCTCTCCAGGCC 717
Db 952 GC-----GCTGGCCCGAGGCCATCGCGCAATCTTCTCACAGCCT 993
Qy 718 ggcctgcgctgcgtgccttttaggaagagcgtgcaggatgggggctccgaagcagcgc 777
Db 994 GGACTGCGCTGCTGCGCTTTCAGGAAGGAGTCCAGGACGCTAGCCTCGCAGTAGCAC 1053
Qy 778 ttctcagctccttcagagagagacattgagaaccacctattagcggacacaattatt 837
Db 1054 TT-----TAGTCTCCTTTGACATGACATAGAGAACCACTCATCGGTGGGCACAATG 1107
Qy 838 gtgcagccacagatatcgaggaataatgaactatgctcttcacdatggccagctgaa 897
Db 1108 GTTCAGCCACACAGATGAGAGGAAACCACTATGCTGTTTCCAGATGGCCCATCTGNA 1167
Qy 898 gtttaacctcatcgtcctgcaccccaaaaaatagcattggagaaaaattttaaggagata 957
Db 1168 GTTTACCTCATCAGTCTGACACCAAAAAGATTGCATGGAGAAAAATTTTAAGGAGATA 1227
Qy 958 tcttttgcctcagggcatcagacacgtggaccactttgggtttatctgtcgggagctct 1017
Db 1228 TCCCTTTGTCTCAGGGCATCAGACATGTGACCACATTTTGGATTTCATCTGCCGAGAGTGC 1287
Qy 1018 tc-----cggagggtgcgcttcatcttctgtctagctgtttcagtgccacaatgag 1071
Db 1288 TCGGTGGCGGCGAGTGGCGCTTTTCATTTTGTCTGTACGTGTTCCAGTGCACAAATGAA 1347
Qy 1072 gctctggttgatgaataatgatgacctgaacggccttcacggtggccgagtgcaag 1131
Db 1348 GCTCTGTTGACGAGATCATGATGACTCTGAAGCAGGCTTTTACGGTAGCTCGGTGTCAG 1407
Qy 1132 cagacagcttaaggccagccagctgtgtgagggtgccccctgcacagcctgcacaag 1191
Db 1408 CAGACGGCTTAAGGCACACGCCAGCTCTGTGAGGGTGCCCTTGCAGGGCTGCAACAAG 1467
Qy 1192 ctctgtgagagatagaggaatgaattcttccaaacaaactagaactgcaaaagcac 1251
Db 1468 CTCTGCGAAGGATAGAGGGATGAATTCATCTAANAACCAATTAGAACTCCAGAGCAC 1527
Qy 1252 ctgacgacattaccaatcagagcagcaggcagctatttttgaagaggttcagaanaattgaga 1311
Db 1528 TTGACCACACTGACCAATCAGAGCAGGCCACCATATTCGAGGAGGTTTTCAGAAATTTGAGA 1587
Qy 1312 ccgagaaatgagcagcagagagaatgaattgatttatttcttcagagatttatgaa 1371
Db 1588 CCAAGAAACGAGCAGCGAGAGATGAATTAATTTATTTCTTTCTGAGGTGCTTATATGAA 1647
Qy 1372 gagaacagaaagacacacatccatattggggagatgaagcagacatcgcagatggcagca 1431
Db 1648 GAGAAGCAAAAAGCAGCACAGCCACACTGGGGGCCCAAGCAGACACTACAGGTGGCAGCA 1707
Qy 1432 gagaatttggaagtgaattaccaccagctgcacactcgtatttaggtagatatgtcgaaa 1491
Db 1708 GAGAAATTTGGGAGTGACCTGCCACACAGTGTAGCGGTTACAGGTTAGATTGCGTGAAG 1767
Qy 1492 acaagcaaaagagattttaacagagcttttagaagatttttcccggggtaataaa 1551
Db 1768 AACAGAGCAAAAGAGTCTTTAACAGAGTCCCTTAGAGAGCAATTCGTGCCGGGGTAATAA 1827
Qy 1552 gccagagggcctgcaggaacactcactcagtggtgattctgtagatcctcctgtctagata 1611
Db 1828 GCCAGAGGCTGCAGGACCAATTCGGCCAGTGTGGATCTGGACAGCTCCACTTCTAGTACT 1887
Qy 1612 ttaagtaaccacagcaagacatctgtgtgtgtaaaaggagccttgcacctctctgag 1671
Db 1888 CTAAGTAACACCAAGAGCTGTCCATGGGTGACAAGAGGCTTCCCGCTCTCTGAG 1947
Qy 1672 agtctcttaagctcctcgtcctcgtgagagacctgtccagtgactcgaagagctatctc 1731
Db 1948 ACCTCTCTCAAGCTCTTGCTCTCAGATGATGACTGTCCAGTACTCAGAGGCCACATTT 2007

Qy 1732 ccagaagcagcagctccctgcgtctgccccagcagggccttcaggaagcagcaaacacctg 1791
Db 2008 GCAGAAGAGTCTGCCCTGTGTGTACCCCGACGAGCGGTTTCAGAAGGAGAGSCCAACCCCTG 2067
Qy 1792 agtctctcccatcgaatgcccaggaaacctccacaacctgccccgggggtccccgggggtt 1851
Db 2068 AGTCAATTTCCCAAGTAGAGTGCCCTGCCCTCCAGAACCTTCCAGAGCTCTCCAGGGGTC 2127
Qy 1852 tcgcaaaagaaacttatgaggtatcactcagtgagcacagacagcctctcatgaagcaag 1911
Db 2128 TCTCAAAGAAACTCATGCGGTACACCTCCGCTGAGCACAGAGACGCTCATGACGCAAG 2187
Qy 1912 gactttgaatccaaagcaaacatcttgtgtattctgtgtgggactccttgtgaagaccgg 1971
Db 2188 GACTTTGAATCCAAAGCAAAACCACTGGGTGACACAGATGGGACCCCTGTAAGACCCGG 2247
Qy 1972 aggccttctcagcagcagatatctcctccagtagccacccccccagcagaagcgtgcgat 2031
Db 2248 CGGCACCTGCTGGAGACAGCAGATATTCCTTCGAGTGGCCACTCCACAGAAAGCTTGTGAC 2307
Qy 2032 tcttcagcagatatgaagattattcagagctgggagagcttccccccagctctctctta 2091
Db 2308 TCCCAGCAGATATGAGATTTATTTCCGAGCTGGGAGAGCTCCCTCCAGCTCCCTTTA 2367
Qy 2092 gaaccagtttgtgaagatggggccctttggccccccacccagcaggaagaaaggacatct 2151
Db 2368 GAACCGTGTGTGAGACGCGCCCATTTGG-CCAGTACAGGAAGAAAGAGAGAGCTCA 2426
Qy 2152 cgtgagctcagagagctgtggcaaaaggctatttctcaacagatactg-ctgcttgaat 2210
Db 2427 CCGGAGCTTCGAGAGCTGTGGAAAAAGGCCATCTTTCAGCAGATCTCGCTCGTAGGAT 2486
Qy 2211 ggagaagaaatcagaagctccaagcctctgaaaatgattgttgaaacaaagcgcctgaa 2270
Db 2487 GAGAGAGGAGATCAGAGCTACAGCCTCTGAAAACGATTTGCTGAAACAAACGCTCAA 2546
Qy 2271 gctgattatgaagaaattactcctcctgtcttaagaagtaactacagtggtgggaaagat 2330
Db 2547 GCTTGACTATGAAGAAATCAGTCCCTGTCTTAAGAAGTCACTACAGTGTGGGAAAGAT 2606
Qy 2331 gcttagcactccagaagatcaaaatgaattgacatgggaaaaatgcactcggtctg 2390
Db 2607 GCTTAGACTCCAGGAAGATCCAAATTAAGTTTGACATGGAAAAAGTGCATCAGCTGT 2666
Qy 2391 tgggcaaggtgtgccacgtcatccacgaggtgaaatcttgaaattctagctgagcaatt 2450
Db 2667 TGGGCAAGCTGTGCCAGCTCATCACCGAGGTGAGATCTGGAAATTTCTAGCTGAGCAGT 2726
Qy 2451 ccaccttaaaacccagtttccagcaaaacagcagcaaaaggatgtgccaatacaagaact 2510
Db 2727 CCACCTTAACCCCATTTCTTAGTAACACAGCAGCCAAAGGAGCTGCCCTTACAAAGCT 2786
Qy 2511 cttaagcagctgacttcccagcagcatgctatttattgacctggggggaacctttcc 2570
Db 2787 CTGAAGAAGCTGACCTCGCAGCAGCAGCCCATCTCTCATCGACCTCGGGGGAACCTTTC 2846
Qy 2571 tacacacccatcttctcctcagcttggagcagcagctatcgtcttaacaatttt 2630
Db 2847 AACACATCCATACTCTCTGCCAGCTTGGAGCAGGTGCTGCTCACTTTTACAACTTCT 2906
Qy 2631 gaagcctactcacttctagaccaggaagtggatatttccaaggtctcagctttgtagc 2690
Db 2907 GAAGGCTACTCGCTTCTGGACAGAGGTTGGATTAATGCTCCAAAGGTCCTCAGCTTTTGGC 2966
Qy 2691 aggcatttgccttctcatatgagtgaggaagagcggttttaaagtctcaagtcttgat 2750
Db 2967 AGGCATTTTGTCTTCTACATGAGTGAGGAGAGGCTTCAAGATGCTCAAGTTCCTGAT 3026
Qy 2751 gtttgacatggggtcggaacagtatcgggccagacatgattattttacagatccagat 2810
Db 3027 GTTTGACATGGGCTCGGGAAACAGTATCGGCCAGACATGATTATTTTGACAGATCCAGAT 3086

Nakamura, Y., Nagahara, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T.
 NEDO human cDNA sequencing project
 Unpublished
 2 (bases 1 to 2362)
 Isogai, T. and Otsuki, T.
 Direct Submission
 Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
 Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan
 (E-mail: genomics@helix.co.jp, Tel: 81-438-52-3951, Fax: 81-438-52-3952)
 NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing;
 Research Association for Biotechnology; cDNA library construction;
 5'- & 3'-end one pass sequencing and clone selection; Helix
 Research Institute (supported by Japan Key Technology Center etc.)
 and Department of virology, Institute of Medical Science,
 University of Tokyo

FEATURES source

Location/Qualifiers
 1..2362
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HEMB1001684"
 /tissue_type="whole embryo, mainly body"
 /clone_lib="HEMB1"
 /dev_stage="embryo, 10 weeks"
 /note="cloning vector: pME18SFL3"
 40..2064
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAB55057.1"
 /db_xref="GI:14041976"
 /translation="MLNKAARSLTESILSRGNKARGLOPHSISVDLSDLSSTL

CDS

SNPSKPSVCEKEALPISESKLIGSSDLSDSESHLPEEPAPLSPQOAFRRANT
 LSHFPECPPQAPGSPGVSQRKLMRYHSVETEPHERKDFESKANHIGDSGGIPV
 KRRHSWRQDIFLRTAPQKACDSSSYEDYSELGELPPRSPLEPCEDGPPFPPEE
 KRTSRELREIWKAIQIILLMEKNOKLAQSENDLKRLKLDYEITPLCKEY
 TTWERMLSTPGSKTKFDMKMHSAVGQVPRHRRGEIWKFLAEQFHLKHQPSKQO
 PKDVPYKELLQLTQSQAHLIDLGRFTFPHVPSAQLGAGQLSNLYLKVASILLQEE
 VGYCOGLSFVAGILLHLMSEEAAPKMLKFLMFDGRLKQYRPMIILLOMYOLSRLL
 HDYRDLNHEEHEGFSLYAPWELTFMFSQPLGFVAFVDFMFLQGETEYFKA
 LSLGSHKPLILQHENLETIVDFIKSTPLNGLVQMEKTIQNVFENDIAKQLQAYEVE
 YHVLQEEILSDLSPLSNORMDKLEKTNSSLRKQNLDLLEQLQVANGRIQSLATIEKL
 LSSSESKLQAMLTLELERSALLQTEVLERRRSRPSDREPECTQPEPTGD"

BASE COUNT 657 a 581 c 578 g 546 t

ORIGIN

Query Match 59.4%; Score 2112; DB 9; Length 2362;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1444. agtgaattaccacccagtcgactcgatttaggtagatgctgtaaaacaaagcaag 1503
 Db 1 AGTGAATTACCACCCAGTCGCACATCGATTTTAGGCTAGATATGCTGAAACAAAGCAAAG 60
 Qy 1504 agatctttaacagagctttgaagattttgtccggggtataaagccagagcctg 1563
 Db 61 AGATCTTTAACAGAGCTTTTAGAAGATATTGTGTCGCCGGGTATTAAGCCAGAGGCGCTG 120
 Qy 1564 caggaacactccatcagtggtgatctgtagactccctgtctgtagtacatgaagtaaacacc 1623
 Db 121 CAGGAACACTCCATCAGTGTGGATCTGGATAGTCCCTGTCTAGTACATTAAGTAACACC 180
 Qy 1624 agcaagagccatctgtgtgtaaaagagggccttgccactctctgtagagctctttaag 1683
 Db 181 AGCAAGAGCCATCTGTGTGTGAAAGAGAGGCGCTTGCCATCTCTGAGAGCTCTTTAAG 240
 Qy 1684 ctctcgctcctcgagagccttcagtgactcgagagctatctccagagagacca 1743
 Db 241 CTCCTCGGCTCCTCGAGGAGCTGTCCAGTCACTCGGAGAGTCACTCCCAAGAGAGCCA 300
 Qy 1744 gctccgctgtgcccccagagccttcagggagggcgagcaaacacccctgagtcacttccc 1803
 Db 301 GCTCCGCTGTGCCCCACAGGCGCTTCAGGAGGCGAGCAACACCCCTGAGTCACTTCCCC 360

Qy 1804 atcgaatgccaggaaacotccaaacctgcccgggggtccccgggggttttcgcaaaagaaa 1863
 Db 361 ATCGAATGCCAGGAACCTCCACAACCTGCCCGGGGGTCCCGGGGGTTTCGCAAAAGAAA 420
 Qy 1864 ctatagagggtatacactcagtgagcacagagacgctcatgaacaaagagaccttgaatccc 1923
 Db 421 CTATAGAGGTATCACTCAGTGAGCACAGACGCGCTCATGAACGAAGGACTTTGAATCC 480
 Qy 1924 aagcaaacaccatcttggtagattcttggtagctcctctgaagaccccgagagcattctctg 1983
 Db 481 AAGCAAAACCATCTTGGTGTATCTCTGTGGGACTCTCTGTGAAGACCCGAGGACATTCCTGG 540
 Qy 1984 aggcagcagatattcctccagtagtagccaccccgagagggcgtgctgattcttccagcaga 2043
 Db 541 AGGCAGCAGATATTCTCCGAGTAGCCACCCCGCAGAGGCGTGGCATCTTTCAGGAGA 600
 Qy 2044 tatgaagattattcagagctgggagagcttccccacagatctccttttagaacaccgtttgt 2103
 Db 601 TATGAAGATTATTCAGAGCTGGGAGAGCTTCCCCCAGCATCTCCTTTAGAACCAAGTTGT 660
 Qy 2104 gaagatgggccccttggccccccaccagagagaaagaaagacacatctctgtagctcga 2163
 Db 661 GAAGATGGGCCCTTTGGCCCCCACCAGAGAGGAAAGAAAGACATCTCGTAGCTCCGA 720
 Qy 2164 gagctgtggcaaaaggctatttctcaacagatctgctgcttagaaggagagaaat 2223
 Db 721 GAGCTGTGGCAAAAGGCTATTCTTCAACAGATACTGCTGTCTTAGAATGGAGAGAAAT 780
 Qy 2224 cagaagctcacaagcctctgaaaaatgatttgcgtaacaagcgcctgaagctcgattatgaa 2283
 Db 781 CAGAAGCTCCAAGCTCTGAAATGATTGCTGAACAAGCGCTGAAGCTGATTATGAA 840
 Qy 2284 gaaattactccctgtcttaagaaagtaactacagctgtgggaaaaagatgcttagcactoca 2343
 Db 841 GAAATTACTCCCTGCTCTTAAAGAAAGTAACACAGTGTGGGAAAAAGATGCTTAGACAC 900
 Qy 2344 ggaagatcaaaaattaaagtttgacatggaaaaaataagcactgctgttgggcagaggtg 2403
 Db 901 GGAAGATCAAAAATAAGTTTGAATGACATGGAATAATGACACTGGCTGTGTGGCAAGGTGTG 960
 Qy 2404 ccagctcatcacccaggtgaaatctgaaatttctagctgagcaatccacacttaaacac 2463
 Db 961 CCAGCTCATCACCGAGGTGAATCTGGAATTTCTAGCTGAGCAATTTCCACCTTTAAACAC 1020
 Qy 2464 cagttccccagcaaacagcagcgaagatgctgccaataaagaaactcttaagacagctg 2523
 Db 1021 CAGTTTCCCAGCAAAACAGCAGCCAAAGGATGTGCCATACAAAGAACTCTTTAAAGCAGCTG 1080
 Qy 2524 actccagcagcagatgagattcttattgaccttggggaacaccttccacacccatcac 2583
 Db 1081 ACTTCCCAGCAGCATGGCATGCTTTATTGACCTTGGGGGAACCTTTCTTACACACCCATAC 1140
 Qy 2584 tctctgccagcttgagcagcagcagctatcgctttacacacttttgaagcgctactca 2643
 Db 1141 TTCTCTGCCAGCTTGGAGCAGGACGACTATCGCTTTTACAACATTTTGAAGGCGCTACTCA 1200
 Qy 2644 ctctagacagcaggaagtgggatatgccaaggtctcagcttggtagcagcagcatttctct 2703
 Db 1201 CTTCTAGACAGGAAGTGGGATATTGCCAAGGTCTCAGCTTTGTAGCAGGCAATTTTCTT 1260
 Qy 2704 ctctaatgagtgaggaagagcgttttaaatgctcaagttctgattgttgacatgggg 2763
 Db 1361 CTTTATATGAGTGAGGAAGAGGCGCTTTAAATGCTCAAGTTTCTGATGTTTGCATGATGGG 1320
 Qy 2764 ctgcggaaaacagtagtcgcccagacatgattattttacagatccagatgtaccagctctcg 2823
 Db 1321 CTGCGGAAACAGATATCGCCAGACATGATTATTTTACAGATCCAGATGTACCAGCTCTCG 1380
 Qy 2824 aggttgcttcattgattaccacagagacacctctacaatacacctggaggagcagagatcggc 2883
 Db 1381 AGGTTGCTTATGATATTACCACAGAGACCTCTACAAATCACTTGCCTGGAGGACGAGATCGG 1440

[illegible]

	Wolfe Street, Baltimore, MD 21205-2185, USA
FEATURES	Location/Qualifiers
source	1..1995 /organism="Bos taurus" /db_xref="taxon:9913" /tissue_type="retinal" /dev_stage="adult" 2..1672 /gene="lyn" <2..1672 /gene="lyn" /codon_start=1 /product="lynceln" /protein_id="CAA76943.1" /db_xref="GI:3893077" translation="ARAGVSRKLVYHVSSTPHERKDFESKADHI\$DA\$ASTPVKT RRHSRQOIFLRVATPKAKSEPRVEDYSELGPPRPSPLEPCVDGGPGFVPPEBK RTSHREIWKALIOQLILLRMKEKNOKLOASENDLLNKRLLDYEEITPCLKEVET WBKILSTPORSKIKDMEKMHSAGCVGPVHRHGIEWFLAEQYLHLKPFFCKQQOPK VTYKELLTQSQAAILDIDJRGTEPTTPPYSAQGLAGOLSYNILKAYSLDDQVG YCQGLSFVAGLILLHHGEFEAFNMFLMFDMLGRKYQIPDMIILOIQIMYSLRLHD YHRDLXNHIEPSIGLSVAAPFLVFASQFPPLGFVARVFMIFLQGSEVIFKVALS VLGSHPKLIILOHNLTIVDFIKFTLPNLGLVMOMEXTISOVFETDITSKOLOAVEVYH LQBELLDSSPLSDNRMDKLENTSSLRKQNLDDLEOLOVANGRISQSL\$EATVEKLTT SESRLKQTALSLERSALLQTVQRRTAELGQSESDFTLPKPSGD"
gene	
CDS	
BASE COUNT	560 a 516 c 497 g 422 t
ORIGIN	
Query Match	38.28; Score 1359.2; DB 4; Length 1995;
Best Local Similarity	88.48; Pred. No. 1.1e+286;
Matches 1475; Conservative	0; Mismatches 193; Indels 0; Gaps 0;
Qy 1844	cggggttgcgaagaaacattatgaggatatcactcagtgcagcacagagacgcctcatg 1903
Dd 9	CAGGGGTCTCGAAAGGAACCTTGAGGTATCACTCAGTCAGCACAGAGACGCCTCATG 68
Qy 1904	aacgaagaagcatttgaaatccaaaagcaaacocatcttggtgattctcggtgggaacctgtga 1963
Dd 69	AACGAAAAGGACTTTGAGTCCAAAGCCGACCACATCAGCGATGCCAGCAGGACCCCTGTGA 128
Qy 1964	agaccggaggcatctcctggaggcagcagatattccctccgagtagccaccgcccgagaag 2023
Dd 129	AGACCGAGACACTCGTGGAGCGCAGCAGATCTTCCTTCGGTGGCCACCCCCCAGAAGG 188
Qy 2024	cgtgcgatcttcacgacagatatgaagattattcagagctgggagagctccccccacat 2083
Dd 189	CCTCGAGTCCCCCAAAAGATACGAAGATTACTCAGAGTTGGGAGAGCTTCCCCECAGAT 248
Qy 2084	ctcctttagaaccagtttgtgaagatgggcccctttggcccccccccacagagaaaaa 2143
Dd 249	CTCCACTAGAACCCAGTTGTGAAGATGACCCTTTGGCCCCGGTCCCGAGAGAAAAGAAAA 308
Qy 2144	ggacatctcgtgagctccgagagctgtggcaaaagggtattcttccaacagatactgctgc 2203
Dd 309	GAACTCTCATGAAC TTCGGGAGCTGTGGCAAAAGGCCATCTACACAAGATATTGCTGC 368
Qy 2204	ttaagaatggaaaggaataatcagaagctccaagcctctgaaaaatgatttgcctaacaagc 2263
Dd 369	TCAGGATGGAAAAGAAAACCAAAACTGCAAGCCTCTGAGAATTGATTTGCTGTAACAAGC 428
Qy 2264	gacctgaagctgattatgaagaaaattactccctgtctttaagaagtaactcacagtgcgg 2323
Dd 429	GCCATAAGCTTGATTATGAAGAAATCACTCCCTGCTTTAAGAAGTAGTACATACAGTGTGGG 488
Qy 2324	aaagatgcttagcacctccaggaagatacaaaaaattaagtttgacatgagaaaaatgcact 2383
Dd 489	AAAGATACTTTAGCACTCCAGGAAGATCAAAATTAAGTTTGACATGGAATAATGCACCT 548
Qy 2384	cggctgttggaagggtgtgccacgtcatcacogaggtgaaatcttgaaatttctagctg 2443
Dd 549	CGGCTGTGGCCAAGGTGTGCGCGACACACACCCGGGTGAAATCTGAAATCTTGAAATCTTGCGCG 608

Db	61	TCCCCCTTTAGAACCGGTGTTGAGACCGCCCATTTGGCCCGCAGTACAGGAAGAAAAAGAGG	120
Qy	2143	aggacatctcgtgagctccgcagagctgtggcaaaaggctattcttccaagatactactgctg	2202
Db	121	AAGACTCACGCCGAGCTTCGAGAGCTGTGGAAAANGGCCATCTTGCAGCAGATGCTTGCTG	180
Qy	2203	cttagaatggagaaggaataacagaagctccaagccttgaaaatgatattgctgaaacaag	2262
Db	181	CTCAGATGGAGAAGGAGAACTCAGAAGCTTACAAGCCTCTGAAAACGATTTCGTGAACAAA	240
Qy	2263	cgcctgaagctcgattaagaaagaattactccctgtctttaaagaagttaactaacagtgtg	2322
Db	241	CGCCTCAAGCTTGACTATGATGAATCACTCCGTGTCTTAAGAAGTCACTACAGTGTGG	300
Qy	2323	gaaaagatgcttagcaactccaggaagatcaaaaaattaaagtttgacatggaaaaaatgcac	2382
Db	301	GAAAAGATGCTTAGCACTCCAGSAGATGCCAATAATTAAGTTTGACATGGAAAAGTGCAC	360
Qy	2383	tcggctgttgggcgaaggtgcccacgcatcacccagagtgaaaatcttgaaattcttagct	2442
Db	361	TCAGCTGTTGGGCAGGTGTGCCACGTCAATCACCAGSGTGAGATCTGGAAATTTCTAGCT	420
Qy	2443	gagcaattccaccttaaaccacagttccccagcaaaacagsgccaaagggatgtgcoatac	2502
Db	421	GAGCAGTTCACCTTTAAACACCCATTTCCTAGTAACAGCAGCAGCCAAAGGACGTGCCCTAC	480
Qy	2503	aagaagaactttaaacagcgtgacttcccagcagcatcgattcttatgaccttgggcga	2562
Db	481	ANAAGCTCTTGAAAGCTGACCTCGCAGCAGCAGCGCATTCATCGACCTCGGGCGGA	540
Qy	2563	accttcttacacacca tacttctctgccagcttgagcaggagacagctatcgctttac	2622
Db	541	ACCTTTCCAACACATCCATTCTCTGCCAGCTTGGCAGCAGTCAAGCTCTCATTTTAC	600
Qy	2623	aacattttgaagsgcctaactactctbtagaccagggaagtgyggatatgtgccaaaggtctcagc	2682
Db	601	AACATTCTGAAGGCCCTACTCGCTTCTTGACCACGAGGAGTGGATACTGCCAAGGTTCTCAGC	660
Qy	2683	tltgtacgagcaetttgtctctcatatgagtgaaggaaggcggtitaaaaagtcctaag	2742
Db	661	TTTGTGGCAGGCCATTTTGTCTCTCACATGAGTAGGGAAGAGGGCGTTCAAGATGCTCAAG	720
Qy	2743	tttctgatgtttgacatggggctgcggaacacagtatcgccagacatgattattttacag	2802
Db	721	TTCTGTATGTTTGACATGGGGCTGGGGAACAGATATCGCCAGACATGATATTTTCGAG	780
Qy	2803	atccadatgtaccagctctcagagttgtctcatgattaccacagagacctctacaatcac	2862
Db	781	ATCCAGATGTACCAGCTGTACGGCTCTCCACGATTTACCACCAGAGACCTCTACAACCCAC	840
Qy	2863	ctggagagcacagagatcggcccagcctctacgtgcccccctggtttctcaccatgttt	2922
Db	841	CTGGAAGAGCACAGATCGGGCCAGCCTCTACGGCGGTCTCCCTGGTTCTCACCGTGTTC	900
Qy	2923	gcctcacagttcccgctgggattcgtagccagagctcttgatagatttttcttcaggga	2982
Db	901	GCCTCACAGTTCCCACTCGGCTTTGTAGCCAGAGTCTTTTGATATGATCTTCTTCAGGGA	960
Qy	2983	acagaggtcataatttaaagtggtctttaagtcgtgttgggaagccataaagcccttgattctg	3042
Db	961	TCAGAGGTCTATATTTAAAGTAGCTTTAAAGCTTTTGGGGGGCCATTAAGGCCCTTGATTCCTA	1020
Qy	3043	cagcatgaanaacctagaacacatagttgactttataaaaaagcagcgtaccaccaaccttggc	3102
Db	1021	CAGCATGAGAACCTGGAAACCATCGGTGGACTTCATTAAGAACAACATCCCACACTGGGC	1080
Qy	3103	tgggtacagatggaagaagaccatcaatcaggtatttgaaatggacatcgtaaacagtta	3162
Db	1081	CTGGTGCAGATGGAGAAGACCATCAGTCAAGTGTGTTTGAGATGGACATCGCCNAGCAGCTC	1140
Qy	3163	caagcttatgaagttgagtaaccacgctcttcaagaagaacttatcgattctctctctc	3222

Db	1141	CAGCCCTATGAGGTCGAGTACCACTGCTCAGGAGGAGCTTATTGAGTCTCTCCCTCTC	1200
Qy	3223	agtgacaaccaaaagatggaataattagagaaaaccacagcagcttacgcgaacagaac	3282
Db	1201	AGTGACACCAACAAATGGAGAAATTTGAGAGAAACCAACACGACGCTTGGCGCAACAGAAC	1260
Qy	3283	cttgacctctcttgaaacagttgcaggtgaggtggcaaatggttaggattccaaagacgttgaggccaac	3342
Db	1261	CTTGACCTCTCTGGAGCAGTTTCAGGTGCGCAATGCTAGGATCCAAAGACCTTTGAAGCCACG	1320
Qy	3343	attgagaagctccctgagcagtgacagcagctgaagcagccatgcttaccttagaactg	3402
Db	1321	GTAGAGAAATCTTACACGCGAGATGAGCTGAAGCAGGCTGCGCTGACCTCGAGGTG	1380
Qy	3403	gagcggtcgccctgctgcagacggtgagagctgcggcgcgagcgcagagccagc	3462
Db	1381	GAGCGCTCCGCGCTGCTGCAGATGGTGAGAGAGCTGGGAGGCAAGCGCCGCCCGCCAGC	1440
Qy	3463	gaccgggagcctgagtgagtcagccagcccgagccccagcgcgagctgacagct	3512
Db	1441	ACTCCAGAGCCAGACTGCACCCAGCTGGAGGCCACAGCGCGATTGACCGCT	1490

RESULT

8

AK057182

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

AK057182

Homo sapiens cDNA FLJ23620 fls, clone STOMA2000386, highly similar to Bos taurus mRNA for linceln.

AK057182.1 GI:16552778

oligo capping; fls (full insert sequence).

Homo sapiens stomach cDNA to mRNA, clone_lib:STOMA2 clone:STOMA2000386.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (sites)

Suzuki,O., Sasaki,N., Aotsuka,S., Shoji,T., Ichihara,T., Shiohata,N., Matsumoto,K., Hirano,M., Sano,S., Nomura,R., Yoshikawa,Y., Matsumura,Y., Moriya,S., Chiba,E., Momiyama,H., Onogawa,S., Kaeriyama,S., Satoh,N., Matsunawa,H., Takahashi,E., Kataoka,R., Kuge,N., Kuroda,A., Satoh,I., Kanata,K., Takami,S., Terashima,Y., Watanabe,M., Sugiyama,T., Irie,K., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuo,Y., Nagai,K. and Isogai,T.

NEDO human cDNA sequencing project

Unpublished

2 (bases 1 to 1788)

Isogai,T., Otsuki,T. and Sugiyama,T.

Direct Submission

Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)

NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

Location/Qualifiers

1..1788

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="STOMA2000386"

/tissue_type="stomach"

/clone_lib="STOMA2"

/note="cloning vector: pME18SFL3"

BASE COUNT ORIGIN	481 a	412 c	434 g	461 t
Query Match	28.1%	Score 998.4;	DB 9;	Length 1788;
Best Local Similarity	99.9%	Pred. No. 1.3e-207;		
Matches	999;	Conservative	0;	Mismatches 1; Indels 0; Gaps 0;
Qy	2556	tggcggaacatttctctacacaccacttctctgcccagcttgagcagagacgctatc	2615	
Db	553	TGGCGGAACCTTCTCTACACACCCTACTCTCTGCCAGCTTGAGCAGCAGCAGCTATC	612	
Qy	2616	gctttacaacattttgaagccttactcaactcttagacagagaagtggatattgcgaagg	2675	
Db	613	GCTTTACAACATTTTGAAGGCTACTCACCTTCTAGACAGGAAGTGGGATATGCCAAGG	672	
Qy	2676	tctcagctttgtagcagcattttgcttcttctcatatgagtgaagagcgctttaaat	2735	
Db	673	TCTCAGCTTTGTAGCAGGCATTTTGTCTTCTCATATCAGTGAGGAAGCGGTTTAAAT	732	
Qy	2736	gctcaagtcttgatgtttgacatgggctgcggaacacagatagggccagacatgattat	2795	
Db	733	GCTCAAGTTTCTGATGTTGACATGGGGCTCGCGAACAACAGTATCGCCAGACATGATTAT	792	
Qy	2796	tttacagatcagatgtaccagctctcaggttcttcattgattaccacagagacctcta	2855	
Db	793	TTTACAGATCAGATGTACACAGCTCTCGAGGTCTCTGAGTGTCTTATGATTACACAGACCTCTA	852	
Qy	2856	caatcacttgagagcagcagatcgccccagcctctacgtcgccccctgggtccctcac	2915	
Db	853	CAATCACCCTGGAGGACGAGATCGCCCCAGCCTCTACGCTGCCCTGGTTCCCTCAC	912	
Qy	2916	catgttgcctacagttcccgctgggattgctagccagagctttgatgattttct	2975	
Db	913	CATGTTTGCCTCACAGTTCGCGTGGGATCGTAGCCAGAGCTCTTTCATATGATTTTCT	972	
Qy	2976	tcagggaacagaggtcatatttaagtggctttaagtctgttggaagccataagccctt	3035	
Db	973	TCAGGGAACAGAGGTCAATTTAAAGTGGCTTTAAGTCTGTGTGGGAAGCCATTAAGCCCTT	1032	
Qy	3036	gattctcagatgaacacattagaaacattagttgactttataaaagcagcgtaccacaa	3095	
Db	1033	GATTCTGCAGCATGAAACCTAGAAACCATAGTTGACTTTTATAAAAGCAGCCTACCCAA	1092	
Qy	3096	ccttggttggtacagatggaagacacacataatcaggtatttgaaatgacatcgctaa	3155	
Db	1093	CTTTGGCTTGGTACAGATGGAAGAGACCATCAATCAGGTATTTGAAATGGACATCGCTAA	1152	
Qy	3156	acagttacaagcttatgaagttgagtcaccagctctctcaagaagaacttatcgattcttc	3215	
Db	1153	ACAGTTACAAGCTTATGAAAGTTGAGTACACAGCTCTTCAAGAAGAACTTATCGATTCTCT	1212	
Qy	3216	tcctctcagtgacacacaaagattggaataattagagaaacacacagcagcttaacgaa	3275	
Db	1213	TCCCTCAGTGACACCAAGATGATGATAAATTAGAACAGACCAACAGCAGCTTACGCAA	1272	
Qy	3276	acagaaccttgacctcttgaaagttgaggttgaggcgaaatggtaggtacaaagccttga	3335	
Db	1273	ACAGAACCCTTGACCTCTTGAAAGTTGACAGTGGGAAATGGTAGGATCCAAAGCCCTGA	1332	
Qy	3336	ggccaccattgagagctctgagcagtgagacgaagctgaagcagagccatgcttacctt	3395	
Db	1333	GGCCACCATGGAAGAGCTCTGACAGTGAAGCAAGCTGAAGCAGGCCATGCTTACTCTT	1392	
Qy	3396	agaactggagcgttcgcccctgctgcagacggttgaggagctgcggcgagcgcagaga	3455	
Db	1393	AGAACTGGAGCGTTCGCCCTGCTGCAGACGCTGAGAGAGTGGCGGCGGAGCGCAGA	1452	
Qy	3456	gccagcgagccggagcctgagtcacagcagcccgagcccgagcagcagcagcagctcg	3515	
Db	1453	GCCAGGACCGGGAGGCTGAGTGCAGCAGCCCGAGCCCGAGCGGAGCTGACAGCTCG	1512	
Qy	3516	caggagagattgcaacacccatccacacactgtccagcctt	3555	

Db 1513 CAGGAGAGATTGCAACACCATCCACACTGTCCAGGCCTT 1552

RESULT 9

AB011175

LOCUS

AB011175

DEFINITION

AB011175

ACCESSION

AB011175.1

VERSION

GI:3043729

KEYWORDS

AB011175.1

SOURCE

Homo sapiens male brain cDNA to mRNA, clone_11b.pBluescriptII SK plus clone.HG1488b.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 5922)

AUTHORS

Ohara, O., Nagase, T. and Ishikawa, K.

TITLE

Direct Submision

JOURNAL

Submitted (13-FEB-1998) Osamu Ohara, Kazusa DNA Research Institute, DNA Technology, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:cdhainfo@kazusa.or.jp, Tel:+81-438-52-3913, Fax:+81-438-52-3914)

REFERENCE

2 (sites)

AUTHORS

Nagase, T., Ishikawa, K., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O.

TITLE

Prediction of the coding sequences of unidentified human genes. IX. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro

JOURNAL

DNA Res. 5 (1), 31-39 (1998)

MEDLINE

98290545

FEATURES

Location/Qualifiers

1. 5922

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="HG1488b"

/sex="male"

/tissue_type="brain"

/clone_lib="pBluescriptII SK plus"

348. 4247

gene

/gene="KIAA0603"

348. 4247

CDS

/gene="KIAA0603"

/codon_start=1

/product="KIAA0603 protein"

/protein_id="BAA2529.1"

/db_xref="GI:3043730"

/translation="MEPPSCIDPEFPHPLEPEPGVSAQPGKPKSKRFLRWYVGGSCLDHRTTLPMLMAEIRRSOKPEAGCGCAPAAREVILVLSAPFLRCVPAPGAGASGTSPTSATOPNPAVFIFEKAOHSREIHNHDLTYFAVLKAPDDPESOMACHYFRATDPSQVPDVISSIRQLSKAMKEDAKPSKDNEDAFYNSOKFEVLKCGVTYTHKAPSLIDDCMEKFLSHEQRLKIQEGRPDGEDADLELVVYVPGSPDCLPBEADGDTHLGLPAGASOPALTSSRVCFPERILEDGFEQOEFRCSSVTVQRRVHBSQSKQPRRHASAPSHQPSDSERNMLFQVGRFEINLISPTKSVVLEKFNKDISCSQGIKVDHFGFCRESPEGLSQICYVFCQASSELVDEMLTKQAFSTAAALQSAKQIKLCEACPMSHLKLCERIEGLYPPRAKLTORHLSLTDNEQADIFERVQKMPVSDQENELVILHRLQCEAKQKTHVHIGEPSTISNSTIPENATSSGREFKLDILKNKRRSLSSLENIERSGANNRGLSGVDSFERNSLASEKDYSPGDSPPGTPPASPPSSAQVTPPEEDSPQFRRRAHTFSHPSTSRKLNLDQRAQVRSPLLRQSSQESQCSNLSVRMYKESNSSSLSTFSAPSTFSLKFSYQNSGRSPQYENIQRDNTASES

SOEGEGRKTSSTCSNESLVSQVSTPPRISWRQRIFLRVASPMKSPSAMQOQDGLD

NOBGLPLSPSTMBEELVIFLSDGDDPEKIEERKSLRWLKAHQITLLLRM

EKENOKLEGASRDELOSKVKLDYEVGACQEVLTMDKLLNCRKLRCDMEDLTH

LUKENGVPKSRGEIQWFLALQYLRHLRHLNKKQPPDISYKELLKQLTAQHMLDVLG

RTFPTHPYSVOLGQGLSLLNKAYSLLDKEVGYCQGISFVAGVILLHMSQEQAFE

MKFLMYDLGFRKQYRDMMSLQIQMYLSRLHLDYRDLNHLNEENISPSLYAAP

FTLFLAQSFSLGFAVRFDIIFLQGVIFKVALSLSSQETLIMECSFENVEFLK

NTPDMMNTSEMEKIIITQVEMDISKOLHAYEVHYVLQDELQESSYSCDESETEKLE

RANSOLKRONMDLEKLVQAHTKIQIALESNENLLTRETIMKMSLIRLTLEQEKWAYQKT

VBQRLKLPADALANCDLLRLDNCNPNKAKGNKP"

BASE COUNT

1624 a 1408 c 1432 g 1458 t

ORIGIN

Query Match	19.2%;	Score 582.2;	DB 9;	Length 5922;
Best Local Similarity	56.6%;	Pred. No. 1.9e-138;		
Matches 1585;	Conservative	0;	Mismatches 1028;	Indels 186; Gaps 10;
Qy 823	agcgagacataatgtgcagccacagatcagagaaatcgaaactatctcttcacg	882		
Db				
Qy 1368	AGGCCCCAGTCAGTCCAGCCCTGGACTCGGAGAAACACAGACCATGCTCTTCAG	1427		
Db				
Qy 883	attggccagtcgaagtttacctcatcagtcctgacacccaaaaatagcattggagaaa	942		
Db				
Qy 1428	GTTGGGCGATTGAGATTAACTTATCAGTCCAGACACTAAATCAGTTGTGCTAGAAAAG	1487		
Db				
Qy 943	aattttaagagatatctttgtcttcagggcaccagacgfggagccactttgggttt	1002		
Db				
Qy 1488	AAATTTAAAGATATCTCTCTGTGTCAGGGTATAAAGCATGTGATCATCTTTGGCTTT	1547		
Db				
Qy 1003	atctgtcgggagtcctccggaggtggcgcttccattttgtctgtacgtgttttcagtcg	1062		
Db				
Qy 1548	ATCTGCCGGAGTCTCCAGAGCCTGGACTTACGCCAGTATATTGTGTATGTTATCCAGTGT	1607		
Db				
Qy 1063	acaaatgaggtctcgtgttgatgaaattatgatgacctgaaacaggcccttcacaggtggcc	1122		
Db				
Qy 1608	GCCAGCAATCTCTGTTGATGAGGTAAATGCTACTCTGAAACAGGCCCTTCAGTACGGCG	1667		
Db				
Qy 1123	gcagtgcagcagacagctaaagcgccagccagctgtgtgagggctgcgcccttcgaaagc	1182		
Db				
Qy 1668	GCTCCCTGCAGATGCCAAGACCCAGATTAACCTGTGTGAGGCCCTGCCGATGCACCTCT	1727		
Db				
Qy 1183	ctgcacagctctgtgagaggatagagggaatgaattcttccaaaaacaaactagaactg	1242		
Db				
Qy 1728	TTGCATTAAGCTCTGTGAAGAGATTGAAGGTCTCTACGCCACCAAGAGGCAAGCTGGTGATA	1787		
Db				
Qy 1243	caaaagcactgcagacattaaaccaatcaggagcagcgactatttttgaagaggttcag	1302		
Db				
Qy 1788	CAGAGGATCTCTCATCTACAGATTAATCAGCAAGCTGACATCTTTGAAAGAGGTTTCAG	1847		
Db				
Qy 1303	aaattgagccgagaaatgagcagcgagagaaatgaattgattttcttctctgagatgt	1362		
Db				
Qy 1848	AAATTAAGCCAGTCAGTGACCCAGGAAGAAATGAACCTGTGTATTTACACCTGAGGCAG	1907		
Db				
Qy 1363	ttatatgaagagaacagaaagaacacatccatttggggagatgaagcagacatgcag	1422		
Db				
Qy 1908	CTGTGTGAAGCCAGCAGACAGACACATGCGTCACATCGGGG-----AAGGCCCTTCT	1958		
Db				
Qy 1423	atggcagcagagaatttggaaagtgaattaccacccagtcgcactcgatttagctagat	1482		
Db				
Qy 1959	ACTATTTCAATAGTACAAATCCAGAAAAATGCAACAGCAGTGGGAGGTTCAAACTTGAC	2018		
Db				
Qy 1483	atgctgaaaaacaaagcaagagatctttaacagagtcttttagaaagtattttgtccogg	1542		
Db				
Qy 2019	ATTCTGAAAAATAAGCTTAAGAGATCTCTTAATAGTCCCTCGGAAAAATATCTTCTCAAGG	2078		
Db				
Qy 1543	gg---taataagccagggcctgcaggaaacctccatcagt-----	1581		
Db				
Qy 2079	GGAGCTAACAAATGAGAGGTGCGCTTGGAGTGTGGACAGTTTTGAACGGTCCACACAGT	2138		
Db				
Qy 1582	-----	1581		
Db	-----			
Qy 2139	CTTGCTTCAGAGAAGGACTACTCACAGGGGATTCTCCACAGGGAGACACCCCGACGCTCC	2198		
Db	-----			
Qy 1582	-----gtgactcgtgatactccctctcagt	1608		
Db				
Qy 2199	CCACGGTCTCAGCTTGGCAAAACGTTTCCCGAAGAGAGATCCGACTCCCCGACGTTTCGA	2258		
Db				
Qy 1609	acattaaagtaacacacagcaagagccatctgtgtgaaagaggccttgcccatctct	1668		
Db				
Qy 2259	AGACGGGCACACAGTTTCAGGCCACCCACCTTCAAGCACAAAGAGAAAGCTGAATTTGCAG	2318		
Db				
Qy 1669	gagagctcctttaagctcctcgtcctcctcg-----gagg	1702		
Db				
Qy 2319	GATGGGAGGGCTCAGGGGTGTCGGTTCCTCTGCTGAGGCAGAGCTCCAGTGAACAGTGC	2378		
Db				

QY	1703	acctgtccagtactcggagagtcattctcccagaagaagaccagctcgcgtgtgcccccaagg	1768
DB	2379	AGCAATCTTTTCGTACGTTGCACGCATGTACAAGAGAGTAATTTCTTCCCTCCAGCTTTCCA	2438
QY	1763	aggcttcaggaggcgacaaacaccccttagtcaacttcccatacgaaatgccagagaacctc	1822
DB	2439	AGTCTTCACACTTCTCTGCGCCCTCCTTTCACCTGCCCCCTCTTTCTTGAAAAGCTTT	2498
QY	1823	cacaacctgccccgggggtc---cccgggggttttcgcaaaggaaaacttatgaggtatacct	1879
DB	2499	TACCAGAATAATTCAGGTAGACTGTCCCACACAGTATGAATAATGAATCAGAACAGACTGCT	2558
QY	1880	cagtgcacacagagagcctcatgaacgaagaagacttgaatccaagaacaaacctcttg	1939
DB	2559	TCAGAACTCAAGTGATGAGGAAGGGGAAAGAGACCTCATCTTACCTGCAGCAANTGAGTCC	2618
QY	1940	gtgatct-gtgtggactccttgaagaccccggaggcattctctgggcagcagatatctc	1998
DB	2619	CTAAGTGTGGGAGGAACCTCTGTCACTCTCGCGGATCTCTCTGGCGCAGCGATTTC	2678
QY	1999	ctcgcagttagcccccgcagaaaggcgtgcgattcttccagcagatatgaagattattca	2058
DB	2679	CTCAGGGTGTCTTCTCCATGAACTATCTCCCTCAGCAATGCAACACAGATGGATGTG	2738
QY	2059	gagctgggagagcttcccccacgactctctttagaacagttgtgaagatggggccctt-	2117
DB	2739	GACAGGAACGAGCTGTGCGCACCTCTCCCCCTCTCTCCACCATGGAGGAGAACCGCTG	2798
QY	2118	-----tgccccccccacagaggagaaaaagaaagaaagacatctcgt	2154
DB	2799	GTTATATCTCTGCTGGGAGGATGACCCAGAAAAAGATTGAAAGAAAGAAAAATCAAAA	2858
QY	2155	gagctccgagagctgtgcacaaaggctattottcaacagatactctgcttagaatggag	2214
DB	2859	GAATCTGAGGAGCTGTGGGAAAGGCTATACACCAACAAATCTTGTACTTCGAATGGAA	2918
QY	2215	aaggaaaaatcagaagctcca- --agcctctgaaaaatgattgtgaaacaagcgcctgaag	2271
DB	2919	AAAGAAAAACCAGAAACTTGAAGGAGCAACGACAGATGMACTCCAGTCCAGAAAAAGTTAAA	2978
QY	2272	ctcgattatgaagaaatactcoctgtcttaagaagaataactacagttgggaaaaagatg	2331
DB	2979	TTAGACTATGAAGAAGTTGGTGCATGTCAGAAGAGGCTTTAATAACTTGGGATAAGAAG	3038
QY	2332	cttagcactccaggaagatcaaaaaatgaatttgacatggaaaaaatcactcgcctggtt	2391
DB	3039	TTGTITAA--ACTGCAGAGCTAAATCAGATGTGATATGGAAGATATTCATACTCTTCTT	3095
QY	2392	gggcaaggtgtccaggtcatcaccaggtgaaatctcggaaattctagctgagcaattc	2451
DB	3096	AAAGAAAGAGTGTCCAAAAGTCCAGCAGAGAGAAATTTGGCAGTTTCTGGCTTTACAGTAC	3155
QY	2452	cacctaaacacacagtttcccagcaaacacgacgacaaaggatgtgccatacaagaacctc	2511
DB	3156	CGACTCAGACACAGATTGCCATTANAACACAGCCCTCGTAGATATCTATAGAAGAACTT	3215
QY	2512	ttaaagcagctgacttcccagcagcatcgatcttattgaccttgggcgaaaccttccct	2571
DB	3216	TTGAAGCAGCTCACTGTCTACGACAGCATGGGATTCCTGTGGATTTAAGGAAGACGCTTCCT	3275
QY	2572	acacaccatactctctgccagcttgagcagacagcattcgctttaacaattttg	2631
DB	3276	ACTCACCTTTACTTTTCAGTTACAGCTTTGGCCAGACAGCTGCTCACTGTTTAACCTCCTG	3335
QY	2632	aaggcctactcaactcttagaccaggaagtgggatatattgccaaggtctcagcttttgtagca	2691
DB	3336	AAAGCCTATTCTTTGCTGGACAAAAGATGGGATACTGTACAGGGGATCAGCTTTGTGGCT	3395
QY	2692	ggcattttgtcttctcataatgaadtgagggaagggttttaaaatgctcaagtttctagt	2751
DB	3396	GGAGTCTGCTTCTGCACTGTGATGAAGAGCAAGGCTTTTGAATGCTGAATTTCTCTATG	3455
QY	2752	tttagcatgaggtgcgcaaacagtatcagccagacatgattkattttacagatccagatg	2811

```

Db 3456 TATGACCTCGGCTTCCGCAACGAGTACAGACCTGACATGATGTCGCTGCAGATTCAAATG 3515
Qy 2812 Taccagctctcaggttcttctcattaccacagagacctctcaaatcaccttgagagag 2871
Db 3516 TACCAGCTGTCCAGGCTCTCTCATGACTATACAGAGATCTCTACATCACCTTGAAGAA 3575
Qy 2872 caccagatcgcccccagcctcagctgcgcccccttggttctcaccatgtttgctcaccag 2931
Db 3576 AATGAAATCAGCCCCAGCTTATATGCTGCCCTGGTTCCTCACATTTGTTGCCCTCTCAG 3635
Qy 2932 ttcccgctggatctcagcagagcttcttatgatatttttcttcagggaacagagctc 2991
Db 3636 TTTTCATTAGGATTTGTAGCCAGAGTTTTTGTATATTTTCTTCAGGGAACCTGAAGTT 3695
Qy 2992 atattaaagtggcttttaagtctgttggaagccataagcccttgattctcagcagtgaa 3051
Db 3696 ATATTCAAGTTGCACCTCAGCTACTGAGCAGCCCAAGAGACACATATATAATGAATGTGAG 3755
Qy 3052 aacctagaaccatagttgacctttataaaagcagcgtaccaccaaccttggtttgtaacag 3111
Db 3756 AGCTTTGAAATATTTGTTGAGTTTCTTAAACACAGCTTACCTGATATGATACCTCTGAA 3815
Qy 3112 atggaagaccatcaatcaggtatttgaatggaatgcgtacgtataacagcttacaagcttat 3171
Db 3816 ATGGAATAAATATTACCCAGGTTTTTGAGATGGATATTTCTAAGCAGTTGCATGCCCTAT 3875
Qy 3172 gaagttgagtcaccagctcttcaagaagaacttatcgattctctctcctcctcagt---gac 3228
Db 3876 GAGGTGGAATATCATGTCTACAGGATGAGCTTCAGGAATCTTCATATCTCTGTGAGGAT 3935
Qy 3229 aaccaaagaattggaataattagaaaaaaccaacagcagcttacgcacaaacagacacttgac 3288
Db 3936 AGTGAACCTTTGGAGAACCTGGAGGGCCCAATAGCCCACTGAAAGACAAACATGAGAC 3995
Qy 3289 ctccctgacagttcaggtgcgcaaatggttaggatcccaagcccttgagggccaccattgag 3348
Db 3996 CTCCTAGAAAATTTACAGGTAGCTCTACTAATAATCCAGGCTTGGGAATCAAACTGGA 4055
Qy 3349 aagctcctgagcagtgagcaagctgaagcagggccatcgcttaccttagtaactggagcgg 3408
Db 4056 AATCTTTTGACGAGAGACCAAAATGAAGTCTTTAATCCGGACCCCTGGAACAAAGAAAA 4115
Qy 3409 tcggccctcgtcgcagcagctgagagagctgcggcgcg 3447
Db 4116 ATGGCTTATCAAAAGACAGTGGAGCAACTCCGGAAGCTG 4154

RESULT 10
AC009595/c standard; DNA; HTG; 189811 BP.
XX AC009595;
XX AC009595;
SV AC009595.3
XX 30-AUG-1999 (Rel. 60, Created)
DT 03-JUL-2000 (Rel. 64, Last updated, Version 4)
DE Homo sapiens chromosome 4 clone RP11-392K14 map 4, WORKING DRAFT SEQUENCE,
DE 28 unordered pieces.
XX HTG; HTGS_DRAFT; HTGS_PHASE1.
XX Homo sapiens (human)
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX [1]
RN 1-189811
RP Birren B., Linton L., Nusbaum C., Lander E.;
RT "Homo sapiens chromosome 4, clone RP11-392K14";
RL Unpublished.

```

```

XX 1-189811
RN Birren B., Linton L., Nusbaum C., Lander E., Allen N., Anderson M.,
RP Baker J., Baldwin J., Barna N., Beckerly R., Benn J., Brown A., Castle A.,
RA Cerny J., Colangelo M., Collins S., Collymore A., Cooke P., Dearellano K.,
RA Depayre E., Devon K., Dewar K., Donelan L., Doyle M., Ferreira P.,
RA Fitzhugh W., Forrest C., Funke R., Gage D., Galagan J., Gardyna S.,
RA Gilbert C., Grant G., Hagos B., Heaford A., Horton L., Howland J.C.,
RA Jones C., Kann L., Karatas A., Lehoczy J., Lieu C., Locke K.,
RA McDonald P., Marquis N., McEwan P., McGurk A., McKernan K., McLaughlin J.,
RA Meldrum J., Molla M., Morris W., Morrow J., Mychaleckyj J., Naylor J.,
RA Niloff M., O'Connor T., O'Donnell P., Pavlin B., Peterson K., Pollara V.,
RA Riley R., Roberts D., Roy A., Severy P., Stange-Thomann N., Stojanovic N.,
RA Stone C., Subramanian A., Tesfaye S., Torruella-Miller I., Vassiliev H.,
RA Vo A., Wagner A., Wheeler J., Wu X., Wyman D., Ye W.J., Zody M.;
;
Submitted (27-AUG-1999) to the EMBL/GenBank/DBJ databases.
Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,
Cambridge, MA 02141, USA
On Jun 23, 2000 this sequence version replaced gi:6006228.
All repeats were identified using RepeatMasker:
CC Spit, A.F.A. & Green, P. (1996-1997)
CC http://ftp.genome.washington.edu/RM/RepeatMasker.html
CC ----- Genome Center
CC Center: Whitehead Institute/ MIT Center for Genome Research
CC Center code: WIBR
CC Web site: http://www-seq.wi.mit.edu
CC Contact: sequence.submissions@genome.wi.mit.edu
CC ----- Project Information
CC Center project name: L1755
CC Center clone name: 392_K14
CC ----- Summary Statistics
CC Sequencing vector: M13; M7815; 100% of reads
CC Chemistry: Dye-primer-amersham; 5% of reads
CC Chemistry: Dye-terminator Big Dye; 95% of reads
CC Assembly program: Phrap; version 0.960731
CC Consensus quality: 164704 bases at least Q40
CC Consensus quality: 175594 bases at least Q30
CC Consensus quality: 181024 bases at least Q20
CC Insert size: 214000; agarose-fp
CC Insert size: 187111; sum-of-contigs
CC Quality coverage: 3.0 in Q20 bases; agarose-fp
CC Quality coverage: 3.5 in Q20 bas.
CC * NOTE: This is a 'working draft' sequence. It currently
CC * consists of 28 contigs. The true order of the pieces
CC * is not known and their order in this sequence record is
CC * arbitrary. Gaps between the contigs are represented as
CC * runs of N, but the exact sizes of the gaps are unknown.
CC * This record will be updated with the finished sequence
CC * as soon as it is available and the accession number will
CC * be preserved.
CC *
CC * 1 451: contig of 451 bp in length
CC * 452 551: gap of 100 bp
CC * 552 2663: contig of 2112 bp in length
CC * 2664 2763: gap of 100 bp
CC * 2764 4124: contig of 1361 bp in length
CC * 4125 4224: gap of 100 bp
CC * 4225 6767: contig of 2543 bp in length
CC * 6768 6867: gap of 100 bp
CC * 6868 8750: contig of 1883 bp in length
CC * 8751 8850: gap of 100 bp
CC * 8851 11272: contig of 2422 bp in length
CC * 11273 11372: gap of 100 bp
CC * 11373 13411: contig of 2039 bp in length
CC * 13412 13511: gap of 100 bp
CC * 13512 16962: contig of 3451 bp in length
CC * 16963 17062: gap of 100 bp
CC * 17063 22544: contig of 5482 bp in length
CC * 22545 22644: gap of 100 bp
CC * 22645 26284: contig of 3640 bp in length
CC * 26285 26384: gap of 100 bp
CC *

```

```
CC * 26385 31223: contig of 4839 bp in length
CC * 31224 31223: gap of 100 bp in length
CC * 31224 36014: contig of 4691 bp in length
CC * 36015 36114: gap of 100 bp
CC * 36115 41939: contig of 5825 bp in length
CC * 41940 42039: gap of 100 bp
CC * 42040 48122: contig of 6083 bp in length
CC * 48123 48222: gap of 100 bp
CC * 48223 52983: contig of 4761 bp in length.
CC * 52984 53083: gap of 100 bp
CC * 53084 58094: contig of 5011 bp in length
CC * 58095 58194: gap of 100 bp
CC * 58195 64606: contig of 6412 bp in length
CC * 64607 64706: gap of 100 bp
CC * 64707 71079: contig of 6373 bp in length
CC * 71080 71179: gap of 100 bp
CC * 71180 79886: contig of 8707 bp in length
CC * 79887 79986: gap of 100 bp
CC * 79987 86155: contig of 6169 bp in length
CC * 86156 86255: gap of 100 bp
CC * 86256 95370: contig of 9115 bp in length
CC * 95371 95470: gap of 100 bp
CC * 95471 103712: contig of 8242 bp in length
CC * 103713 103812: gap of 100 bp
CC * 103813 113565: contig of 9753 bp in length
CC * 113566 113665: gap of 100 bp
CC * 113666 123867: contig of 10202 bp in length
CC * 123868 123967: gap of 100 bp
CC * 123968 132296: contig of 8329 bp in length
CC * 132297 132396: gap of 100 bp
CC * 132397 141110: contig of 8714 bp in length
CC * 141111 141210: gap of 100 bp
CC * 141211 161622: contig of 20412 bp in length
CC * 161623 161722: gap of 100 bp
CC * 161723 189811: contig of 28089 bp in length.
XX
FH Key Location/Qualifiers
FH source 1. .189811
FH /chromosome="4"
FH /db_xref="taxon:9606"
FH /organism="Homo sapiens"
FH /map="4"
FH /clone="RP11-392K14"
FH /clone_lib="RP11-392K14"
FH misc_feature 1. .451
FH /note="assembly_fragment clone_end:T7 vector_side:right"
FH misc_feature 532. .2663
FH /note="assembly_fragment"
FH misc_feature 2764. .4124
FH /note="assembly_fragment"
FH misc_feature 4225. .6767
FH /note="assembly_fragment"
FH misc_feature 6868. .8750
FH /note="assembly_fragment"
FH misc_feature 8851. .11272
FH /note="assembly_fragment"
FH misc_feature 11373. .13411
FH /note="assembly_fragment"
FH misc_feature 13512. .16962
FH /note="assembly_fragment"
FH misc_feature 17063. .22544
FH /note="assembly_fragment"
FH misc_feature 22645. .26284
FH /note="assembly_fragment"
FH misc_feature 26385. .31223
FH /note="assembly_fragment"
FH misc_feature 31324. .36014
FH /note="assembly_fragment"
FH misc_feature 36115. .41939
FH /note="assembly_fragment"
FH misc_feature 42040. .48122
FH /note="assembly_fragment"
```

```
FT misc_feature 48223. .52983
FT /note="assembly_fragment clone_end:SP6 vector_side:left"
FT misc_feature 53084. .58094
FT /note="assembly_fragment"
FT misc_feature 58195. .64606
FT /note="assembly_fragment"
FT misc_feature 64707. .71079
FT /note="assembly_fragment"
FT misc_feature 71180. .79886
FT /note="assembly_fragment"
FT misc_feature 79987. .86155
FT /note="assembly_fragment"
FT misc_feature 86256. .95370
FT /note="assembly_fragment"
FT misc_feature 95471. .103712
FT /note="assembly_fragment"
FT misc_feature 103813. .113565
FT /note="assembly_fragment"
FT misc_feature 113666. .123867
FT /note="assembly_fragment"
FT misc_feature 123968. .132296
FT /note="assembly_fragment"
FT misc_feature 132397. .141110
FT /note="assembly_fragment"
FT misc_feature 141211. .161622
FT /note="assembly_fragment"
FT misc_feature 161723. .189811
FT /note="assembly_fragment"
XX
SQ Sequence 189811 BP; 52455 A; 40839 C; 40321 G; 53485 T; 2711 other;

Query Match 13.1%; Score 465.8; DB 30; Length 189811;
Best Local Similarity 99.6%; Pred. No. 5.2e-91;
Matches 467; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 416 aadgctgagatcatcagctccatccgtcagcgccggggaagatcgccgcagagagc 475
DB 97313 AGGTGCTGAGATCATCAGCTCCATCCGTGAGCGGGGAAGATCGCCGCGAGGAGC 97254
QY 476 tgcactgcccgtccgaggttcgacgacacgttttccaaagattcgaggtgtctttcg 535
DB 97253 TGCACGTGCCCGTCCGAGTTTCGACGACACGTTTCCAAAGATTTCGAGGTCTCTTCTGCG 97194
QY 536 gcccgtgacgtggtgcgcacaagaagctccgcgcctgacgcagagtcgacgcaga 595
DB 97193 GCGCGCTGACGTTGGCGCACAGAAGCTCCGCGGCCCTGATCGACGAGTGCATCGAGA 97134
QY 596 agttcaatcacgtcagcgagccggttcgagagagcccccgcagagcccccaccccccacatg 655
DB 97133 AGTTCAATCAGTTCAGCGGCGCGGGGTCCGAGAGCGCCCGCCGCAACCCGCCCATG 97074
QY 656 cccgcccacagagccagagagcctgtgcgagggccatgcgcaagtctcttccagc 715
DB 97073 CCGCGCCACAGGAGCCAGGAGCCTGTGCGCAGGGCCCATGCGCAAGTCTCTTCCAGC 97014
QY 716 ccggcctgcgtcgtggttccttaggaagagctgcagagtggtgggctccgagcagcg 775
DB 97013 CCGCGCTGCGCTCGCTGGCCTTTAGGAGAGAGCTGAGGATGGGGCTCCGAAGCAGCG 96954
QY 776 gcttcttcagctccttcgagagagcgacattgagaacacacctattagcggaacaata 835
DB 96953 GCTTCTTCAGTCTCTTCGAGGAGGCGACATTGAGAACCACCTCATTTAGCGGACACAATA 96894
QY 836 ttgtgcagcccacagatatcgaggaaaatcgaaactatgctcttcacgat 884
DB 96893 TTGTGAGCCCCACAGATATCGAGGAAAATCGAACTATGCTCTTTCACGGT 96845

RESULT 11
AC021106 AC021106 195108 bp DNA linear PRI 03-JUL-2001
LOCUS Homo sapiens clone RP11-177C12, complete sequence.
DEFINITION
```

```
AC021106 AC021106.6 GI:14589687
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 195108)
AUTHORS
Waterston,R.H.
TITLE
The sequence of Homo sapiens clone
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 195108)
AUTHORS
Waterston,R.H.
TITLE
Direct Submission
JOURNAL
Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
3 (bases 1 to 195108)
AUTHORS
Waterston,R.H.
TITLE
Direct Submission
JOURNAL
Submitted (03-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
On Jul 3, 2001 this sequence version replaced gi:14018126.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0177C12
----- Location/Qualifiers -----
1. 195108
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-177C12"
BASE COUNT 53372 a 40112 c 41873 g 59751 t
ORIGIN

Query Match 13.1%; Score 465.8; DB 9; Length 195108;
Best Local Similarity 99.6%; Pred. No. 5.2e-91;
Matches 467; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 416 aagtgcctgagatcagtcacatcgctccatcgctcagcgagggaagatcgcccgagaggagc 475
Db 180759 AGGTGCGCTGAGATCATCAGCTCCATCGTCCGCGGGGAAGATCGCCGCGAGGAGC 180818

Qy 476 tgcactgcccgcagcttcgacacacgctttcccaagaattcgaggtctcttcgcg 535
Db 180819 TGCACTGCCCGTCGAGTTCGACGACGCTTTTCCAGAAAGTTCGAGGTGCTCTTCGCG 180878

Qy 536 gccgcgtgacgtggcgacaaaggctccgcgcgcctgacgcagtgatgcacgaga 595
Db 180879 GCCGCGTGACGGTGGCGCACAAAGGCTCCGCGGCCCTGATCGAGAGTGATCGAGA 180938

Qy 596 agttcaatacgtcagcgagcgcgggggtccgagagccccccccccccccatg 655
Db 180939 AGTTCAATACGTGACGCGAGCGGGGGTCCGAGAGCCCCCGCCCAACCCGCCCATG 180998

Qy 656 ccgcgccacagagcagagcctgtgcgagggcccatgcgaagctcttcccaagc 715
Db 180999 CCGCGCCACAGGAGCGCAGAGCCCTGTGCGAGGCCCATGCGCAAGTCTCTTCCGAGC 181058

Qy 716 ccggcctgcgctcgctggccttaggaagagctgcaggatgggggctccgaagcagcg 775
Db 181059 CCGGCTGCGCTCGCTGCGCTTTAGGAAGAGAGTGCAGGATGGGGGCTCCGAAGCAGCG 181118

Qy 776 gcttcttcagctcttcgagagagacacattagaaaccctcatttagcggacacaata 835
Db 181119 GCTTCTTTCAGCTCTTTCGAGGAGCGACATTTAGAACCCACTCATTAGCGGACACAATA 181178

Qy 836 ttgtcagccacagatcagagaaatcgaactatgctcttcacgat 884
Db 181179 TTGTGAGCCACAGATATCGAGAAATCGAACTATGCTCTTCACGGT 181227

AC108933 208318 bp DNA linear HTG 01-FEB-2002
Homo sapiens chromosome 4 clone RP11-392K14, WORKING DRAFT
SEQUENCE, 5 unordered pieces.
AC108933 AC009595
AC108933.1 GI:18464316
HTG: HTGS_PHASE1; HTGS_DRAFT.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 208318)
AUTHORS
Waterston,R.H.
TITLE
The sequence of Homo sapiens clone
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 208318)
AUTHORS
Waterston,R.H.
TITLE
Direct Submission
JOURNAL
Submitted (01-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
On Feb 1, 2002 this sequence version replaced gi:8671945.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: H_NH0392K14
Drafting center: WIBR
----- Summary Statistics -----
Sequencing vector: M13; #
Chemistry: Dye-terminator Big Dye; # of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 205652 bases at least Q40
Consensus quality: 206608 bases at least Q30
Consensus quality: 207022 bases at least Q20
Insert size: 218000; agarose-fp
Insert size: 207918; sum-of-contigs
Quality coverage: 6.45 in Q20 bases; agarose-fp
Quality coverage: 6.62 in Q20 bases; sum-of-contigs
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 25746: contig of 25746 bp in length
* 25747 25846: gap of unknown length
* 25847 65131: contig of 39285 bp in length
* 65132 65231: gap of unknown length
* 65232 105028: contig of 39797 bp in length
* 105029 105128: gap of unknown length
* 105129 159307: contig of 54179 bp in length
* 159308 159407: gap of unknown length
* 159408 208318: contig of 48911 bp in length.
Location/Qualifiers
1. 208318
/organism="Homo sapiens"
/db_xref="taxon:9606"
FEATURES
source
```

```
/chromosome="4"
/clone="Rp11-392K14"
misc_feature 1..25746
  /note="assembly_name:Contig16"
misc_feature 25847..65131
  /note="assembly_name:Contig17"
misc_feature 65232..105028
  /note="assembly_name:Contig18"
misc_feature 105129..159307
  /note="assembly_name:Contig19"
misc_feature 159408..208318
  /note="assembly_name:Contig20"
BASE COUNT 59195 a 45373 c 44728 g 58621 t 401 others
ORIGIN

Query Match 13.1%; Score 465.8; DB 2; Length 208318;
Best Local Similarity 99.6%; Pred. No. 5.2e-91;
Matches 467; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 416 aagtcctgagatcatcagctccatccgtcagcgcggaagatccgccgagagagc 475
DB 129553 AGTGTGCTGAGATCATCAGTCTCCGTCAGCGCGGGGAGATCGCCCGGAGGAGGC 129494

QY 476 tgcaactgcccgcaggttcgcagacacggtttccaagaagtccaggtctctctcg 535
DB 129493 TGCACCTGCCCGTCCGAGTTCGACGACACAGCTTTTCCAAGAAGTTCGAGGTGCTTCTGCG 129434

QY 536 gcgcgctgacgtgctgcacagaaggctccgcgcgcctgcacagatgcacgcaga 595
DB 129433 GCGCGGTGACGCTGCGCGCAACAAGAGGCTCCGCGCGCCCTGATGACGAGTGCATCGAGA 129374

QY 596 agttcaatcacgtcagcgacgcgggggtccgagagccccccgccccacccgccccatg 655
DB 129373 AGTTCAATCACGTACGCGCACGCGGGGGTCCGAGAGCCCCCGCCCAACCCGCCCCCATG 129314

QY 656 ccgcgccccacaggagcaggtgctgcgcagggcccatgcgcaggtccatgcctctccagc 715
DB 129313 CCGCGCCCCACAGGACCCAGGAGCGCTGTGCGCAGGCGCCATGCGCAAGTCTTCTCCAGC 129254

QY 716 ccgcgctgacgtcgtgctgaggaagagcagtcagagtgaggcgccctccgaacagcg 775
DB 129253 CCGGCGCTGCGCTGCTGCGCTTTAGGAAGAGTCCAGATGGGCGCCCTCCGAAGCAGCG 129194

QY 776 gcttttcagctcttcgagagagcagcattgagaaccacctcattagcgacacaata 835
DB 129193 GCTTCTTACGCTCTTCGAGGAGAGCGACATTGAGAACCACTCATAGCGGACACAATA 129134

QY 836 ttgtgcagcccacagatcagcggaataatcgaaactatgctcttcacgat 884
DB 129133 TTGTGCGAGCCACAGATATCGAGGAAATCGAACTATGCTCTTACGGT 129085

RESULT 13
AL390202_04/c
WPCOMMENT
Sequence split into 10 fragments LOCUS AL390202 Accession AL390202
Fragment Name Begin End
AL390202_00 1 110000
AL390202_01 100001 210000
AL390202_02 200001 310000
AL390202_03 300001 410000
AL390202_04 400001 510000
AL390202_05 500001 610000
AL390202_06 600001 710000
AL390202_07 700001 810000
AL390202_08 800001 910000
AL390202_09 900001 988176
Continuation (5 of 10) of AL390202 from base 400001 (AL390202 Homo sapiens chromosome 20

Query Match 11.9%; Score 424.2; DB 2; Length 110000;
Best Local Similarity 96.0%; Pred. No. 6.1e-82;

Matches 435; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 atgaacaataataacattcacagaggaacacatcgtcttccttaacgaggtctcgtggtgat 60
DB 10365 ATGGAACCAATAACATTCACAGCAAGGAACAATCTGCTCTCTACGAGGCTCTCGTGGAT 10306

QY 61 ttggcctgcagctggtggtccctgcctgctgctgcttccctgacacccatgccccatgctg 120
DB 10305 TTTGGCCTGCAGCTGCTGGGCTCCCTGCTGTGCTATTCCTGACCACCATGCCATGCTG 10246

QY 121 cctctgggttgctgagtgctggaagactcagcagggcagtcacacagaaagaacctgta 180
DB 10245 CCTCGGTTGTGGCTGAGGTGCGAAGACTCAGCAGGCACTCCACAGAAAGAACCTGTA 10186

QY 181 accaagaagtcctgctgttccacctctgacctgagctgagatgtgaacctgagccaggg 240
DB 10185 ACCAAGCAAGTCCGGCTTTGCGTTTCAACCTCTGAGCTGAGATGTGAACCTGAGCCAGG 10126

QY 241 agaagtcacagtggtggtccctgcctgcttattccagcatctttgagtgcaagcctcagct 300
DB 10125 AGAAGTCAACAGTGGGATCCCTGATCTATTCCAGCATCTTTGAGTGCAGGCTCAGCGT 10066

QY 301 gtccacaactgattcacacaagtcacacagtcaccccaagttacttctgctgctgtaaggaa 360
DB 10065 GTTCACAACACTGATTCACAACAGTCAATGACCCCAAGTTACTTTGCTGTGATTAGGAA 10006

QY 361 gacgctgtccaccgagcagatctgctatgtgttcaagccgagatgatacaacaaagt 420
DB 10005 GACGCTGTCCACCGCAGAGTATCTGTATGTGTTCAAGCCGATGATCAACAAAGTA 9946

QY 421 cctgagatcactcagctccatccctcagcgcgggg 453
DB 9945 AGTGAGTGGAGATCCAAAGAGACTAAGGTGTGG 9913

RESULT 14
DMU50542 3214 bp mRNA linear INV 12-JUN-1996
LOCUS Drosophila melanogaster pollux (plx) mRNA, complete cds.
DEFINITION U50542
ACCESSION U50542
VERSION U50542.1 GI:1373162
KEYWORDS fruit fly strain-Oregon R.
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
REFERENCE 1 (bases 1 to 3214)
AUTHORS Zhang,S.D., Kassisi,J., Olde,B., Mellerick,D.M. and Odenwald,W.F.
TITLE Pollux, a novel Drosophila adhesion molecule, belongs to a family of proteins expressed in plants, yeast, nematodes, and man
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3214)
AUTHORS Zhang,S.D.
TITLE Direct Submission
JOURNAL Submitted (04-MAR-1996) Shang-Ding Zhang, Neurogenetics, LNC, NINDS, NIH, 9000 Rockville Pike, Bethesda, MD 20892, USA
FEATURES
source
1..3214
/organism="Drosophila melanogaster"
/strain="Oregon R"
/db_xref="taxon:7227"
/chromosome="3"
/map="83c"
/tissue_type="trachea; CNS"
1..3214
/gene="plx"
130..2328
/gene="plx"
/function="adhesion molecule"
/codon_start=1
/product="pollux"
/protein_id="AAB02200.1"
gene
CDS
```

```
/db_xref="GI:1373163"
/translation="MKPKRGRKDAELRELHRLTAIROTIMLRNMTENAMLAQARON
ENELKRIKLDYIEIVCDKQOLIERNEQIERNSTQIGNKADPKVLGHAIKTVPRSKR
GDVWTELAEOHSWNTAPVDTPREFNPYHMLLKHLLTEHHAIFIDLGRFPNHFQFY
KDPGLGQSLFNLKAYSLIDPGLYCQGLGFCVGLLLHDCDEANSFQLLKHLMPFR
NMRTKYLPMKFKQLQYQLSRVLKDLPLDLYVWLQNDVSPPLYAAPLTLTVFSQF
PLGFVAPVDPKLLFLESDDVTFKAIALLSVHKQOLLAKDNFEIMDYLYKTVPKMEHT
CMFOIMKLVFSDIGKQLAIEYVYVLOBEITVTHHLEMLNREXTQNHLEEQLOLF
AQSSIAQLETRRSQQAQITTLQSQVOSLELTQTLQRYVQGLVERHNDLELNEVR
MLQOLLDQRKRIFTERKIGKSVSNHSLFPLKVLDELFERDELSPQKQKEK
TFPELROOQKHLNGQSSNGVSGSPPTSRPNRLLDNARSARTVMQVLDLKL
LPBHVKQVANIIRKSLVDSVGTPLSPSPSTASNSGGGSIFFSMGYRTTTPALSLAQ
ROSGYVAITTPACPOHMEVEVAPATMAVMOEDVEEPOPMPHPLSMVGGDYNVREKGT
QLKSIPVHHMRAIPLGGVQHPSTPEVAVRVPVPELAPATATATGRS"
BASE COUNT      862 a 798 c 774 g 780 t
ORIGIN

Query Match      8.6% Score 304.6; DB 3; Length 3214;
Best Local Similarity 56.0% Pred. No. 6.1e-56;
Matches 621; Conservative 0; Mismatches 479; Indels 9; Gaps 2;

Qy 2155 gagctccgagagtggtgcaaaaggctattcttcaacagacagactgctgcttagaagtgag 2214
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 169 GAGCTGAGGGAACGTGCGCGACTGCCATTAGGCAGACATATAATGCTGAACCGCATGGAG 228

Qy 2215 aaggaatacagaagctccagcctctgaaaatgattgtgtaacaaagcgctgaagctc 2274
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 229 ACGGAGAACGCGCATGTTGCAGAGCAGCGCAGAAATGAGACGAGCTAAAGCGCATTAACATG 288

Qy 2275 gattataaagaatactccctctctaaagaataactacagtggtgggaaagatgctt 2334
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 289 GACTACGAGGAGATTGATCCCTGGCAGCAAGCAGCTAATCGAGCGATGGGACGACATCAT 348

Qy 2335 ---agcactcaggaagatcaaaataaagtgttgacatggaaaaaatgcactggctgtt 2391
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 349 GAGCGCAACTCACACAGATAGCAACAAGAGGATCCCAAGGCTTGGGCGACGCCATT 408

Qy 2392 gggcaaggtgtgcacgtcatcacccaggtgaaatctcgaaattcttagctgagcaattc 2451
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 409 CGCACCGGAGTACCGCATCGAAGCGCGCGATGCTGGACCTTCTTGGCTGGACGACAT 468

Qy 2452 caccttaa-----acaccagttccacagcaaacagcagcaaaaggatgtgccatacaa 2505
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 469 TCCATGAACACGCGACCGGTGGACACAAGAGCGATTCCCAACTTCATACACCGTATCAC 528

Qy 2506 gaactcttaagcagctgactctccagcagcatcgattcttattgaccttggcgcaacc 2565
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 529 ATGCTGCTGAACACACTTACTAGCATCAGCATCGGATGATTTTCAATTGATCTGGCAGGACT 588

Qy 2566 ttctctacacaccatactctctgccagcttgagcagcagacagctagctcttacaac 2625
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 589 TTTCTTAATACACAGTCTTACAAAGATCCCGTGGTCTCGGCCAGTGTGCTGCTTTTAAAC 648

Qy 2626 atttgaagggctactactcttagaccaggaagtggaatattgccaaaggtctcagcttt 2685
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 649 CTGCTGAAGGGCTACTCATTTAGATCCGGAATTAGGGTACTGCCAGGGTCTGGGCTTC 708

Qy 2686 gtagcagcatatttgcctcttcataatagtgaggaagaggcggtttaaaatctcaagttt 2745
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 709 ATCTGCGGCGCTTACTCTCTGCTGATTCGATGAAGCCAATTCATTTCAACTGCTGAAGCAC 768

Qy 2746 ctgatgtttgacatggggctcggaacagatcgccagcagacatgattattttacagatc 2805
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 769 CTAATGTTTCGCGGTAATATATCGCACGAAATACCTTGGCGGACATGAATAAGTTTCAACTG 828

Qy 2806 cagatgtaccagctctgaggtgttctcatgattaccacagagacctctacaatcacctg 2865
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 829 CAGCTGTATCAGCTCTCCGATTTGGTCAAGGATCATCTGCCAGATCTCTACGTGGGCTC 888

Qy 2866 gagagacacagagatggcccccagcctctacgtcgcccccgtgttcttccaccatgtttgcc 2925
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 889 GATCAGAACGATGTCTCGCCCATTTTGTATGCGGCTCCCATGATGATCTCACCGCTCTTAGC 948
```

```
Qy 2926 tcacagttccgctgggattcgtagccagagctctttagatgatttttttcaggggaaca 2985
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 949 TCACAGTTTCGCGTGGGATTGTCGCCGCGCTCTTTGATCTGCTCTTCTCGAATCTCTCC 1008

Qy 2986 gagtcataattaaagtggcttttaagtctgttgggaagccataagccctttagcttcgag 3045
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1009 GATGTAATCTTTAGTTTGGCATTTGCCCTGTCTCCGTGTCACAGCAACTACTGTGCC 1068

Qy 3046 catgaaacacctagaaacacatagttgactttataaaagcagcgtaccacacacttgcttg 3105
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1069 AAGGATAACTTCGAAGAGATTATGAGCATTTATGAGACCGTGTGTCGCAAAAGATGGAGCAC 1128

Qy 3106 gtacagatgaaagaccatcaatcaggtatttgaatggagacatcgctaaacagcttaca 3165
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1129 ACGTGCATGGAGCAAAATATGAAGCTGGTCTTTAGCATGGACATTTGGAACAGCTTGCC 1188

Qy 3166 gcttatgaagttagtaccacgtctctcaagaagaacttatcgattcctctctctcagtt 3225
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1189 GAATACAACTGGGAGTACAAATGTCTTCAGGAGGAGATTACCACCACTAACCATCACCTA 1248

Qy 3226 gacaacaaagaatggataaattagagaa 3254
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1249 GAAATGCTTAACCGGAGAGACGACGAGAA 1277
```

```
RESULT 15
DME17919
LOCUS      Drosophila melanogaster mRNA for pollux protein. 4746 bp mRNA linear INV 18-NOV-1998
DEFINITION      Drosophila melanogaster mRNA for pollux protein.
ACCESSION      Y17919
VERSION      Y17919.1 GI:3893102
KEYWORDS      pollux gene.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
              Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
              Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 4746)
AUTHORS      Xu,X.Z., Web,P.D., Chen,H., Li,H.S., Yu,M., Morgan,S., Liu,Y. and
              Montell,C.
TITLE      Retinal targets for calmodulin include proteins implicated in
              synaptic transmission
JOURNAL      J. Biol. Chem. 273 (47), 31297-31307 (1998)
MEDLINE      99030403
REFERENCE      2 (bases 1 to 4746)
AUTHORS      Montell,C.
TITLE      Direct Submission
JOURNAL      Submitted (11-AUG-1998) C. Montell, The Johns Hopkins University
              School of Medicine, Department of Biological Chemistry, 725 N.
              Wolfe Street, Baltimore, MD 21205-2185, USA
```

```
FEATURES             source
source
source
gene
CDS
/gene="pollux"
/dev_stage="adult"
/tissue_type="retinal"
/gene="pollux"
/gene="pollux"
/feature="calmodulin binding protein"
/codon_start=1
/product="Pollux protein"
/protein_id="CAA76939.1"
/db_xref="GI:3893103"
/translaton="MAELMHQMRDPATHLGSGVCSIPQTLIGGGGGSHGNSGNALNGI
HATPATNLKMSAMRNAQHDTSANPVSCKMKASKSYTHGLSSSGTNYIPTSTSAQSN
LSLLADISPNTHTFFEYMYVGVKIRVSKRPVNTFIDDLALPKFKAYDAQRLRLQNRKM
```



```
QY 1258 acattaaccaatcagagcagcgactattttgaagaggttcagaaattgagaccgaga 1317
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1368 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1309
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1318 aatgagcagcagagaaatgatttatttctctgagatgttatatgaagagaaa 1377
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1308 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1249
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1378 cagaagaacacatccatttgggagatgaagcagacatcgagatggcagcagagaat 1437
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1248 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1189
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1438 attggaagtgaattaccacccagtcgactgatttaggtagatattgtgaaacaaca 1497
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1188 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1129
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1498 gcaaaagagatcttaacagagcttttagaaagtattttgtccggggtaataaagccaga 1557
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1128 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1069
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1558 ggcctgcaggaacatccatccag 1580
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1068 RRATCGCAAGCTCCTCGACCTG 1046
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
US-08-920-812-13
; Sequence 13, Application US/08920812
; Patent No. 5763188
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,812
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9515 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
```

```
; STRAIN: Clinical Isolate P2-2
; US-08-920-812-13

Query Match 1.3%; Score 45.4; DB 1; Length 9515;
Best Local Similarity 46.2%; Pred. No. 0.033;
Matches 151; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

QY 438 catccgtcagcggggaagatcgcccgcgagcaggagcgtgcactgccgtccgagttcga 497
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 802 CATGGCAAGCCGGTATGGAGCGCTGGAACTGATGTACCGGGCGCCGCCACGCTT 861
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 498 cgacacgttttccaagaagtctcgtctcttcttcgcccgcggtgacggtgagcaca 557
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 862 CGCCTGGTATGCGGAAGCCTCGACAAGCTCTACGACCAAGTCTCGCCGCGCCGCCAGCA 921
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 558 gaaggtcgcgcgcgcctgatcgacgagatgcatcgagaagttaacatcacgtcagcgcag 617
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 922 GACCTTGGCCACCATTACCCTGCGCTGCCGTGGGGGTGATCGGCGCGGTGTGCCGTGGAA 981
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 618 ccggggggtcgcgagagcccgcccaacccgcgcccatgcccgcacacagggagcagga 677
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 982 CTTCCGCTCGACATGGCCGCTTGGAGCTGCCCCGCGCCTGGCCGCGGCAACTCGGT 1041
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 678 gctgtgcgcagggccatgcgaagtctcttcttcacagcccgccgtcgtcgtggtcctt 737
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1042 GGTGCTCAAGCGCGCGAGCAGTCGCCGTCTCCGCCCTGCGCCTGGCGAGTGGCCCT 1101
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 738 taggaagcagctgcagatgggggct 764
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1102 GGAGCGGGGGTGGCGGAAGGGGTGCT 1128
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
US-08-920-827-13
; Sequence 13, Application US/08920827
; Patent No. 5770375
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,827
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9515 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
```

LENGTH: 9515 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Pseudomonas aeruginosa
STRAIN: Clinical Isolate P2-2
US-08-920-827-13

Query Match 1.38; Score 45.4; DB 1; Length 9515;
Best Local Similarity 46.28; Pred. No. 0.033; Indels 0; Gaps 0;
Matches 151; Conservative 0; Mismatches 176;
Qy 438 catcgtcaggcgggaagatcgcggcaggagagtgctcactgcccgtccgagttcga 497
Db 802 CATGGCAAGCGGTGATGGAGCCTGGAAATCATGATACCGCGCGCCCGCAGTCTT 861
Qy 498 cgacacgtttccaaagaagttcgggtgctcttctgcccgcgctgacggtgagcaca 557
Db 862 CGCCTGGTATCGGAAAGCCTCGAACAGCTCTACGACAGGTGCGCGCGCGCCAGCA 921
Qy 558 gaaggtcgcggccctgatcgacagtgatcgatcgagaagttcaatcagtcagggcag 617
Db 922 GACCCGTGCCACCATTAACCCGCGTGGCGTGGGGGTGATCGCGCGGTGGTGGCGT 981
Qy 618 ccgggggttcgagagcccccccaaccccgcccatgcccgcgcccacagggagccagga 677
Db 982 CTTCCCGCTCGACATGCGCGCTGGAGCTCGCCCGCGCCCTGCGCGCGCAACTCGT 1041
Qy 678 gctgtgcccagggccctgacgaagtccttctccagcccgcgctgagctgagcctt 737
Db 1042 GGTGCTCAAGCGCGCGAGCAGTTCGCGCTGCGCCCTGCGCTGCGCGAGCTGGCCCT 1101
Qy 738 taggaagagctgcaggtgagggcct 764
Db 1102 GGAGCGGGGTGCGGGAAGCGGTGCT 1128

RESULT 5
US-08-921-177-13
; Sequence 13, Application US/08921177
; Patent No. 5798211
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921.177
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362.577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547

REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 9515 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Pseudomonas aeruginosa
STRAIN: Clinical Isolate P2-2
US-08-921-177-13

Query Match 1.38; Score 45.4; DB 1; Length 9515;
Best Local Similarity 46.28; Pred. No. 0.033; Indels 0; Gaps 0;
Matches 151; Conservative 0; Mismatches 176;
Qy 438 catcgtcaggcgggaagatcgcggcaggagagtgctcactgcccgtccgagttcga 497
Db 802 CATGGCAAGCGGTGATGGAGCCTGGAAATCATGATACCGCGCGCCCGCAGTCTT 861
Qy 498 cgacacgtttccaaagaagttcgggtgctcttctgcccgcgctgacggtgagcaca 557
Db 862 CGCCTGGTATCGGAAAGCCTCGAACAGCTCTACGACAGGTGCGCGCGCGCCAGCA 921
Qy 558 gaaggtcgcggccctgatcgacagtgatcgatcgagaagttcaatcagtcagggcag 617
Db 922 GACCCGTGCCACCATTAACCCGCGTGGCGTGGGGGTGATCGCGCGGTGGTGGCGT 981
Qy 618 ccgggggttcgagagcccccccaaccccgcccatgcccgcgcccacagggagccagga 677
Db 982 CTTCCCGCTCGACATGCGCGCTGGAGCTCGCCCGCGCCCTGCGCGCGCAACTCGT 1041
Qy 678 gctgtgcccagggccctgacgaagtccttctccagcccgcgctgagctgagcctt 737
Db 1042 GGTGCTCAAGCGCGCGAGCAGTTCGCGCTGCGCCCTGCGCTGCGCGAGCTGGCCCT 1101
Qy 738 taggaagagctgcaggtgagggcct 764
Db 1102 GGAGCGGGGTGCGGGAAGCGGTGCT 1128

RESULT 6
US-08-362-577C-13
; Sequence 13, Application US/08362577C
; Patent No. 5807673
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362.577C
; FILING DATE: 27-MAR-1995

```
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9515 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
; STRAIN: Clinical Isolate P2-2
; US-08-362-577C-13

Query Match          1.3%; Score 45.4; DB 1; Length 9515;
Best Local Similarity 46.2%; Pred. No. 0.033;
Matches 151; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

QY 438 catccgtcagcgagggaagatcgccgcgagagagtgctccttctgcccgcgagtgccgagtgcca 497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 802 CATGGCAAGCGCGGTGATGGAGCGCTGGAACTGATGATGATGATGATGATGATGATGATGATGATGAT 861

QY 498 cgacacgttttccaagaagttcgagtgctccttctgcccgcgagtgccgagtgccgagtgcca 557
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 862 CGCCTGGTATGCGGAAGCGCTCGACAAGCTCTACGACAGGTCTGCGCGCGCGCGCCAGCA 921

QY 558 gaaggtctcgcgccctgatcgacagtgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg 617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 922 GACCTTGGCCACCATTTACCCGCGTCCCGTGGGGGTGATCGCGCGGTGGTGGCGGTGGTGGCGAA 981

QY 618 ccg9999gttcgagagcccgcccaaccgcccacatgcgcgcgcacacagggagccagga 677
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 982 CTTCCGCTCGACATGGCGCGCTGGAGCTCGCCCGCGCGCTGGCGCGCGCGCAACTCGGT 1041

QY 678 gctgtgagcagcccatcgcaagtctcttctccagcccgccgctgctgctgctgctgctgctgctgct 737
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1042 GTGCTCAAGCGCGCGAGCAGTCGCGCTTCTCCGCCCTGCGCGCTGCGCGCTGCGCGCTGCGCGCT 1101

QY 738 taggaagagctgcagatgggggct 764
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1102 GGAGCGGGGGTGGCGGAAGGCGTGCT 1128

RESULT 7
US-08-920-828-13
; Sequence 13, Application US/08920828
; Patent No. 585398
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,828
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9515 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
; STRAIN: Clinical Isolate P2-2
; US-08-920-828-13

Query Match          1.3%; Score 45.4; DB 2; Length 9515;
Best Local Similarity 46.2%; Pred. No. 0.033;
Matches 151; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

QY 438 catccgtcagcgagggaagatcgccgcgagagagtgctccttctgcccgcgagtgccgagtgcca 497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 802 CATGGCAAGCGCGGTGATGGAGCGCTGGAACTGATGATGATGATGATGATGATGATGATGATGATGAT 861

QY 498 cgacacgttttccaagaagttcgagtgctccttctgcccgcgagtgccgagtgccgagtgcca 557
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 862 CGCCTGGTATGCGGAAGCGCTCGACAAGCTCTACGACAGGTCTGCGCGCGCGCGCCAGCA 921

QY 558 gaaggtctcgcgccctgatcgacagtgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg 617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 922 GACCTTGGCCACCATTTACCCGCGTCCCGTGGGGGTGATCGCGCGGTGGTGGCGGTGGTGGCGAA 981

QY 618 ccg9999gttcgagagcccgcccaaccgcccacatgcgcgcgcacacagggagccagga 677
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 982 CTTCCGCTCGACATGGCGCGCTGGAGCTCGCCCGCGCGCTGGCGCGCGCGCAACTCGGT 1041

QY 678 gctgtgagcagcccatcgcaagtctcttctccagcccgccgctgctgctgctgctgctgctgctgct 737
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1042 GTGCTCAAGCGCGCGAGCAGTCGCGCTTCTCCGCCCTGCGCGCTGCGCGCTGCGCGCTGCGCGCT 1101

QY 738 taggaagagctgcagatgggggct 764
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1102 GGAGCGGGGGTGGCGGAAGGCGTGCT 1128

RESULT 8
US-08-253-155A-9
; Sequence 9, Application US/08253155A
; Patent No. 5691147
; GENERAL INFORMATION:
; APPLICANT: Gyuris, Jeno
; APPLICANT: Draetta, Giulio
; TITLE OF INVENTION: CDK4 Binding Proteins
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
```

ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/253,155A
FILING DATE: 02-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 361709
REFERENCE/DOCKET NUMBER: MII-028
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 8201 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-253-155A-9

Query Match 1.3% Score 45; DB 1; Length 8201;
Best Local Similarity 53.0% Pred. No. 0.039;
Matches 96; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 2549 ttgacctggcgaaactttctcacacccatactctctgcccagcttgagcaggac 2608
DB 2133 TGGACGTGAGGACGACTCTCCGGAACCATGCTCTTTAGGATCGATATGGGCCAAGC 2192
QY 2609 agctatcgtttacaacattttgaaggcctactcactctctagaccaggaagtgggatatt 2668
DB 2193 AGAGGNACTATTACATCTCTCTGGCTATTTCGGAGTATACCCGGAGGTGGGCTACT 2252
QY 2659 gccaaagttcagctttgtagcaggcaatttgccttctcatatagtgagaagagcggt 2728
DB 2253 GCAGGAGCTGAGCACATACCGCGCTTGTCTCTCTTTTCTGCTGCTGAGGAGCGCAT 2312
QY 2729 t 2729
DB 2313 T 2313

RESULT 9
US-08-295-060-3
Sequence 3, Application US/08295060
Patent No. 5659123
GENERAL INFORMATION:
APPLICANT: VAN RIE, Jeroen
APPLICANT: JANSSENS, Stefan
APPLICANT: PERPEROEN, Marnix
TITLE OF INVENTION: NEW DIABROTICA TOXINS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/295,060
FILING DATE: 26-AUG-1994

CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Crane-Feury, Sharon E
REGISTRATION NUMBER: 36,113
REFERENCE/DOCKET NUMBER: 010830-052
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1957 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: YES
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 13..1947
US-08-295-060-3

Query Match 1.2% Score 41; DB 1; Length 1957;
Best Local Similarity 47.9% Pred. No. 0.22;
Matches 113; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 488 ccgagttcgacgacagctttcccaagaagttcgaggtgctcttctgcggcggtgacgg 547
DB 1322 CCCAGACCTAGCAGACGNNCGCAACGTGGCGCTGTGAGCTGGACAGCATCGACCAGC 1381
QY 548 tggcgcaacaagaagctccgcccctgatcgacagtgatcgagagtgatgagagttcaatcacg 607
DB 1382 TGCCTCCGAGACCCACCGACGAGCCCTGGAGAGGGGTACAGCCAGCTGAAGTACG 1441
QY 608 tcagcggcagcgggggtccgagagagccccgcacacccccccatgccggccccacag 667
DB 1442 TGATGTGCTTCTGATGAGGCGAGGCGGACCATCCAGTGTGACCTGGACCCACA 1501
QY 668 ggaagccaggagcctgtgcgagggcccatgcgcaagtctcttccagccccggcctg 723
DB 1502 AGAGCGTGACTTCTTCAACATGATCGACAGCAAGATCACCAGCTGCCCTG 1557

RESULT 10
US-08-858-003-2
Sequence 2, Application US/08858003
Patent No. 6060234
GENERAL INFORMATION:
APPLICANT: Katz, Leonard
APPLICANT: Stassi, Diane L.
APPLICANT: Summers Jr., Richard G.
APPLICANT: Ruan, Xiaolan
APPLICANT: Pereda-Lopez, Ana
APPLICANT: Kakavas, Stephan J.
TITLE OF INVENTION: NOVEL POLYPEPTIDE DERIVATIVES
TITLE OF INVENTION: AND RECOMBINANT METHODS FOR MAKING SAME
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Rd.
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,003
FILING DATE: 16-MAY-1979

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dianne Casuto
REGISTRATION NUMBER: P-40,943
REFERENCE/DOCKET NUMBER: 4952.US.P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (847)-938-3137
TELEFAX: (847)-938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1030 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-858-003-2

Query Match 1.1%; Score 40.8; DB 3; Length 1030;
Best Local Similarity 50.0%; Pred. No. 0.18;
Matches 102; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
QY 440 tccgtcagcggggaagatcgcccgagagagctgcactgcccgtccgagttcgacg 499
DB 203 TCACACAGCGCGCTGTTCGCGGTGGAGACCTCCCTGTTCCGGCTGTCGAGGCCACG 262
QY 500 acacgttttccaagaagttcgaggtgctcttctgcccgcgctgacggtggcgacacaaga 559
DB 263 GCCTCTGCCCGACTACCTCATCGGCCACTCCATCGCGGAAGTGACCGGCCACCTGG 322
QY 560 aggtctccgcggccctgatcgagagtgatcgagagtgatcaatcagtcagcgagcc 619
DB 323 CCGGGTCTCGATCTGGCGGAGCGCTGCTGCTCGCCACCGCGCGCCGCTGATGC 382
QY 620 ggggggtccgagagccccccca 643
DB 383 AGTCGCGCGCGCGCGCGCGGA 406

RESULT 11
US-09-078-166-2
Sequence 2, Application US/09078166
Patent No. 6063561
GENERAL INFORMATION:
APPLICANT: Katz, Leonard
APPLICANT: Stassi, Diane L.
APPLICANT: Summers Jr., Richard G.
APPLICANT: Ruan, Xiaocan
APPLICANT: Pereda-Lopez, Ana
APPLICANT: Kakavas, Stephan J.
TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Rd.
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/078,166
FILING DATE: 16-MAY-1979
CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dianne Casuto
REGISTRATION NUMBER: P-40,943
REFERENCE/DOCKET NUMBER: 4952.US.P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (847)-938-3137
TELEFAX: (847)-938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1030 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-078-166-2

Query Match 1.1%; Score 40.8; DB 3; Length 1030;
Best Local Similarity 50.0%; Pred. No. 0.18;
Matches 102; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
QY 440 tccgtcagcggggaagatcgcccgagagagctgcactgcccgtccgagttcgacg 499
DB 203 TCACACAGCGCGCTGTTCGCGGTGGAGACCTCCCTGTTCCGGCTGTCGAGGCCACG 262
QY 500 acacgttttccaagaagttcgaggtgctcttctgcccgcgctgacggtggcgacacaaga 559
DB 263 GCCTCTGCCCGACTACCTCATCGGCCACTCCATCGCGGAAGTGACCGGCCACCTGG 322
QY 560 aggtctccgcggccctgatcgagagtgatcgagagtgatcaatcagtcagcgagcc 619
DB 323 CCGGGTCTCGATCTGGCGGAGCGCTGCTGCTCGCCACCGCGCGCGCTGATGC 382
QY 620 ggggggtccgagagccccccca 643
DB 383 AGTCGCGCGCGCGCGCGCGGA 406

RESULT 12
US-08-997-467-2
Sequence 2, Application US/08997467
Patent No. 6200813
GENERAL INFORMATION:
APPLICANT: Katz, Leonard
APPLICANT: Stassi, Diane L.
APPLICANT: Summers Jr., Richard G.
APPLICANT: Ruan, Xiaocan
APPLICANT: Pereda-Lopez, Ana
APPLICANT: Kakavas, Stephan J.
TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Rd.
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/997,467
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: 08/858,003
APPLICATION NUMBER: 08/858,003
FILING DATE: 16-MAY-1997

ATTORNEY/AGENT INFORMATION:
NAME: Dianne Casuto
REGISTRATION NUMBER: P-40,943
REFERENCE/DOCKET NUMBER: 4952-US-P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (847)-938-3137
TELEFAX: (847)-938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1030 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-997-467-2

Query Match	1.1%	Score 40.8;	DB 4;	Length 1030;
Best Local Similarity	50.0%;	Pred. No. 0.18;		
Matches 102;	Conservative 0;	Mismatches 102;	Indels 0;	Gaps
QY 440	tccgtcagcggggaagatcgcgcgcagagagagctgcactgcgcgtccaggtctcgacg	499		
DB 203	TCACAGGGGCTCTGTCGGCGTGAGACCTCCTGTTCGGCTGTTCGAGGCCACG	262		
QY 500	acagcttttccaaagaattcagagtgctcttctcgcgccgcgtgcagcgtggcgacaaaga	559		
DB 263	GCCTCGTCCCGCACTACCTCATCGGCCACTCCATCGGCAGAGTGACCGCGGCCACCTGG	322		
QY 560	aggctccgcggccctgatccagcagtgatcatcgaagttcaatcacgtcacggcgacgc	619		
DB 323	CCGGGTCTCTGATTTTCGGGGACGCGCTGCTGTGTCCCCACCGCGCCCTGTATGC	382		
QY 620	gggggtccgagagcccccgccca	643		
DB 383	AGTCGGCCCGGGCGCGCGCGGA	406		

RESULT 13

US-08-814-052-19

: Sequence 19, Application US/08814052

: Patent No. 6015783

: GENERAL INFORMATION:

: APPLICANT: von der Osten, Claus

: APPLICANT: Cherry, Joel R.

: APPLICANT: Bjornvad, Mads E.

: APPLICANT: Vind, Jesper

: APPLICANT: Rasmussen, Michael Dolberg

: TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING

: TITLE OF INVENTION: OR STAINS FROM CELLULOSE FABRIC

: NUMBER OF SEQUENCES: 55

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: No. 6015783o No. 6015783disk of No. 6015783th America, Inc.

: STREET: 405 Lexington Avenue, Suite 6400

: CITY: New York

: STATE: New York

: COUNTRY: U.S.A.

: ZIP: 10174-6401

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Diskette

: COMPUTER: IBM Compatible

: OPERATING SYSTEM: DOS

: SOFTWARE: Fastseq for Windows Version 2.0

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/814,052

: FILING DATE: 06-MAR-1997

: CLASSIFICATION: 510

: ATTORNEY/AGENT INFORMATION:

: NAME: Lambiris, Elias J

: REGISTRATION NUMBER: 33,728

: REFERENCE/DOCKET NUMBER: 4684.204-US

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2249 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-814-052-19

	Query Match	1.1%	Score 40.6	DB 3	Length 2249
	Best Local Similarity	48.9%	Prod. NO. 0.31		
	Matches 109	Conservative	0	Mismatches 114	Indels 0
	Gaps				
Qy	474	gctgcactgccgtccgagttcgcagcacgctttcccaagaagttcgcaggtgctctctcg	533		
Db	1234	GCTCTGGCGGGCTCCAGGAATCCCTACCGCGCGCCATCTTCCACTACGCGGGCGCCC	1293		
Qy	534	cggcgcgctgcaggtgtggcgacaaagaagctccgcggccctgatcgacgagtgcatacga	593		
Db	1294	CGGCGGCGCGGCCACGAGCAGAGGGCAAGGCCCGGTTCGACCACTGGCTGGACCTCC	1353		
Qy	594	gaagttaatcacgtccagcggcgagccgggggtccgagagcccccgcaccaaccgcgcaca	653		
Db	1354	CAACCTAAGCCCGTGTGGCCCGCGACGTGCCCTTGACGGGTTCGCCAAGGGCGCGA	1413		
Qy	654	tgcgcgccccacaggagccaggagcctgtgtgcgagggcccatg	696		
Db	1414	CAACACGCTCGAGTCACTCCCTCGACACACACGGGCGACGCCCTTG	1456		

RESULT 14
US-08-814-052-17
Sequence 17, Application US/08814052
Patent No. 6015783
GENERAL INFORMATION:
APPLICANT: von der Osten, Claus
APPLICANT: Cherry, Joel R.
APPLICANT: Bjornvad, Mads E.
APPLICANT: Vind, Jesper
APPLICANT: Rasmussen, Michael Doiberg
TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING
TITLE OF INVENTION: OR STAINS FROM CELLULOSEIC FABRIC
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: NO. 6015783o No. 6015783disk of No. 6015783th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U. S. A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/814,052
FILING DATE: 06-MAR-1997
CLASSIFICATION: 510
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/POCKET NUMBER: 4684,204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2279 base pairs
TYPE: nucleic acid

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2002, 14:07:12 ; Search time 5191.84 Seconds
(without alignments)
9241.748 Million cell updates/sec

Title: US-09-762-311-3_COPY_171_3725

Perfect score: 3535
Sequence: 1 atgggaacaaataacattcac.....tccacactgtccaggccctt 3555

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:.*
1: em_estba:.*
2: em_esthum:.*
3: em_estin:.*
4: em_estmu:.*
5: em_estov:.*
6: em_estpl:.*
7: em_estro:.*
8: em_htc:.*
9: gb_est1:.*
10: gb_est2:.*
11: gb_htc:.*
12: gb_gss:.*
13: em_gss_hum:.*
14: em_gss_inv:.*
15: em_gss_pin:.*
16: em_gss_vrt:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
c	1	833.6	23.4	1001	10 278359 HS278359 Hu
	2	826.6	23.3	1051	10 BM479578 AGENCOURT
	3	819.2	23.0	1007	10 BM016445 603641252
	4	782.4	22.0	784	10 BG828092
	5	741.8	20.9	835	10 BG759286 602710975
	6	738.8	20.8	800	10 BG746377 602703644
	7	737.4	20.7	1121	10 BM476629 AGENCOURT
	8	729.8	20.5	1109	10 BM460573 AGENCOURT
	9	722.2	20.3	952	10 BE299948 600944494
	10	713.8	20.1	838	10 BI907513 603065640
	11	705	19.8	807	10 BG120498 6027046815
	12	701	19.7	832	10 BM014940 603640877
	13	689.8	19.4	851	9 AL570425 603640877
	14	683.4	19.2	754	10 BE300005 600944594
	15	682.6	19.2	892	10 BE561889 601346094
	16	678.6	19.1	686	10 BG769758 602744658
	17	675	19.0	1108	10 BG326728 602425604

ALIGNMENTS

RESULT 1

278359/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

278359 1001 bp mRNA linear EST 28-JUL-1999
HS278359 Human fetal brain S. Meier-Ewert Homo sapiens cDNA clone
3.114 (CEPH) 5', mRNA sequence.
278359 1 GI:1495132
EST.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1001)
Meier-Ewert, S., Lebre, A.S., Holbert, S., Saada, C., Bougueleret
J., Massart, C., Guillou, S., Gervy, P., Poullier, E., Rigault, P.,
Weissenbach, J., Lennon, G., Chumakov, I., Dausset, J., Lehrach, H.,
Cohen, D. and Cann, H.M.
Survey of CAG/CTG repeats in human cDNAs representing new genes:
candidates for inherited neurological disorders
Hum. Mol. Genet. 5 (7), 1001-1009 (1996)
96414310
Contact: Neri C.
Fondation Jean Dausset - CEPH
27 Rue Juliette Dodu, 75010 Paris, France
Related sequence: 278360
5'-sequence (upper strand).
Location/Qualifiers
1.1001
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="3.114 (CEPH)"
/clone_lib="Human fetal brain S. Meier-Ewert"
/tissue_type="brain"
/dev_stage="fetus"
/note="cDNA library of S. Meier-Ewert, Max Planck
Inst.f.Mol.Genetics, Berlin, FRG"
237 a 229 c 231 g 272 t 32 others

18 671.8 18.9 978 10 BG029927
19 669.6 18.8 670 9 AL557623
20 669.4 18.8 857 9 AL544177
21 668.4 18.8 750 10 BI084159
22 664.6 18.7 795 10 BE560184
23 661 18.6 661 9 AL046787
24 661 18.6 948 10 BG258116
25 659.2 18.5 672 10 BE675153
26 652.8 18.4 806 9 AW043925
27 648.8 18.3 665 9 AW173375
28 647.8 18.2 812 10 BI758489
29 641 18.0 1161 10 BM466841
30 637.8 17.9 773 10 BE361312
31 636.4 17.9 862 10 BI910369
32 635.4 17.9 660 10 BE276812
33 632.4 17.8 761 10 BF797592
34 632 17.8 753 10 BF305442
35 630 17.7 722 10 BE560392
36 628 17.7 748 10 BG424359
37 624.4 17.6 627 9 AW772402
38 621.2 17.5 626 9 AW239183
39 612 17.2 623 10 BE884440
40 604.4 17.0 649 10 BE513375
41 597 16.8 605 9 AW732976
42 596.2 16.8 622 10 BE514108
43 592.4 16.7 774 10 BI763670
44 590.2 16.6 952 10 BG424757
45 589.4 16.6 758 10 BG166636

BG029927 602297271
AL557623 AL557623
AL544177 AL544177
BI084159 602869636
BE560184 601346896
AL046787 DKFZPA34M
BG258116 602379212
BE675153 7F03D07.X
AW043925 WY82B07.X
AW173375 X178Q07.X
BI758489 603023334
BM466841 AGENCOURT
BE361312 601344450
BI910369 603068025
BE276812 601178843
BF797592 602257390
BF305442 601892937
BE560392 601347108
BG424359 602447963
AW772402 hn73a08.x
AW239183 xb37a10.Y
BE884440 601510724
BE513375 601315702
AW732976 bb18b11.Y
BE514108 601315945
BI763670 603049558
BG424757 602453405
BG166636 602339021

Query Match 23.4%; Score 833.6; DB 10; Length 1001;
Best Local Similarity 93.5%; Pred. No. 2.2e-197;
Matches 937; Conservative 0; Mismatches 56; Indels 9; Gaps 8;

QY 2146 acatctgtgagctccgagagctgtggcaaaaggctattcttcaacagatactgtgttt 2205
|||||
DB 1001 ACATCTCGTGTGAGCTCCGAGAGCTGTGCTTAAAGGCTATTTCTTCAACAGATNCTGTGCTN 942
|||||
QY 2206 aqaatggagaagaaacacagaagctcccaagcctctgaaatgattgtctgaacagcgc 2265
|||||
DB 941 AGATGGAGAGAAATCAGAGCTCCAGGCTCTGAAATGATTTGCTGAACAGCGC 882
|||||
QY 2266 ctgaagctgattatgaagaataactccctgtcttaagaagaagtaactacagtggtgaa 2325
|||||
DB 881 CTGAAGCTCGATTATGAAGAAATNACTCCCTGTCTTAAAGAGTAACTACAGTGTGGAA 822
|||||
QY 2326 aagatgcttagcactccagaagat--caaaaaataagtttgacatggaataatgcact 2383
|||||
DB 821 AAGATGCTTAGCAGCTCCAGGAAGATCCCAANAAATTAAGTTTGACATGGAAATAATGCAT 762
|||||
QY 2384 cggctgttgggc-aaggtgtgcacgtcatccagaggtgaaatctggaaa-ttctagc 2441
|||||
DB 761 CGACTGTGGGTAGGTGTGCCAGCTCATCACCAGGTGAATCTGGAAATTTCTAGC 702
|||||
QY 2442 tgagcaa-ttccaccttaaacaccagtttcccagcaaacacagcagcaaaaggtgtgcac 2500
|||||
DB 701 TGAGCAATTTCCACCTAAACACACAGTTTCCAGCANACAGCAGCCANAGGATGTGCCAT 642
|||||
QY 2501 acaagaactcttaagaagctgaacttcccagcagcatcgattcttattgaccttggc 2560
|||||
DB 641 ACAAGAACTCTTAAAGACAGCTGACTTCCAGCAGCATCGGATTCATTGTGACCTTGGNC 582
|||||
QY 2561 gaaccttctctacacaccatacttctgtccagctgttgagcagcagcagctatgcctt 2620
|||||
DB 581 GAACCTTTCTACACACCCATCTTCTGTGCCAGCTTGGAGCAGCAGCATGCTGCTTT 522
|||||
QY 2621 acaaatgttgagcctactcaactctagacaggaagtggatattccaaagttctca 2680
|||||
DB 521 ACAACATTTTGAAGGCTACTCATCTTAGACAGGAAGTGGATATTGCCAAGGTCTCA 462
|||||
QY 2681 gctttgacagcatttgccttcttcatatgaggaagagcgttttaaatgctca 2740
|||||
DB 461 GCTTTGACAGGCAATTTGCGGCTTCATATGATGAGGAAGAGGCGGTGAAGATGCTCA 402
|||||
QY 2741 agttctgtattgtgacatgggctgcggaacagatcgccagacatgattattttac 2800
|||||
DB 401 AGTTTCTGATG-TTGACATGGGCTNNGGNACAGNATCGGCCAGACATGATATTATTAC 343
|||||
QY 2801 agatccagatgtaccagctctcgaggtgttcttcattgattaccacagagaccttacaatc 2860
|||||
DB 342 AGANCCAGATGTACCACTCTCGGGGTTNCNTNATGATTTACCACAGAGACCTCAACAATC 283
|||||
QY 2861 acctgagagacagagatcgcccccagcctctacgctg-ccccctggttctccacctg 2919
|||||
DB 282 ACCTGAGGAGGNCAGATCGGCCCCAGGCTCTACGCTGCCCCCTCGTGTCTCTCACCATG 223
|||||
QY 2920 ttgctctcagcttccgctgggattcgtagccagagcttttgatgatattttcttcag 2979
|||||
DB 222 TTTGCTCAGCTGTCGCTGGGATTCGTATNCAGAGNCTTTGANAATGATTTTCTTCAG 163
|||||
QY 2980 ggaacagaggtcatatttaagtggtcttaagctgttgggaagccataagcccttgatt 3039
|||||
DB 162 GGAACAGAGTCTATATTTAAAGTGGCGGNAAGNCTGNTGGGAANCCATAAGCCNGTGAT 103
|||||
QY 3040 ctgacagatgaaacacctagaaa--ccatagttgactttataaaagcagccttcccaacct 3098
|||||
DB 102 CTCANCATGAAGAACCTAGAAAGCCAGATNGACTTATGTAAAGCAGCAGCTACCCCAACCT 43
|||||
QY 3099 tggcttggtacagatgaa-aagaccatcaatcaggtatttg 3139
|||||
DB 42 TGCGGTGTTNAGATGGAAGNAGACCATCAATCAGGTAGTG 1

RESULT 2

BM479578 BM479578 1051 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT_6464998 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5577183
DEFINITION 5', mRNA sequence.
ACCESSION BM479578
VERSION BM479578.1 GI:18528620
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1051)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM12330 row: d column: 16
High quality sequence stop: 662.

FEATURES
source 1..1051
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5577183"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 296 a 278 c 234 g 243 t
ORIGIN

Query Match 23.3%; Score 826.6; DB 10; Length 1051;
Best Local Similarity 95.8%; Pred. No. 1.3e-195;
Matches 892; Conservative 0; Mismatches 34; Indels 5; Gaps 4;

QY 2451 ccaccttaaacaccagtttcccagcaaacacagcagcagcaaggatgtgccatacaagaact 2510
|||||
DB 1 CCACCTTAAACACCAGTTTCCAGCAACAGCAGCAGCAAGAGGATGTGCCATACAAAGAACT 60
|||||
QY 2511 cttaaagcagctgacttcccagcagcagcttcttattgaccttggcgcaaccttcc 2570
|||||
DB 61 CTTAAAGCAGCTGACTTCCAGCAGCATCGGATTCATTGACCTTGGCGGAACCTTTC 120
|||||
QY 2571 tacacacccatactctctgccagcttggagcagcagcagctatcgctttacaacattt 2630
|||||
DB 121 TACACACCATACTCTCTGCCAGCTTGGAGCAGCAGCATCGCTTTTACAACTTT 180
|||||
QY 2631 gaagcctactcaactctctagacacgaagtgggatattcccaaggtctcagctttgtacg 2690
|||||
DB 181 GAAGCCTACTACTCTTAGACAGCAAGTGGGATATTGCCAAGGTCTCAGCTTTGTAGC 240
|||||
QY 2691 agcatttgccttctctcatatgaggaagcggcgtttaaatactcaagttctgat 2750
|||||
DB 241 AGCATTTTGTCTCTTCATATGAGTGAGGAGAGGCGTTTAAATGCTCAAGTTTCTGAT 300
|||||
QY 2751 gtttgacatgggctgcggaacacagtatcgccagacatgattattttacagatccagat 2810
|||||
DB 301 GTTTCACATGGGCTCGGGAACACAGTATCGCCAGACATGATTATTTTACAGATCCAGAT 360
|||||


```
/db_xref="taxon:9606"  
/clone="IMAGE:4851090"  
/clone_lib="NIH_MGC_48"  
/tissue_type="primary B-cells from tonsils (cell line)"  
/lab_host="DH10B (phage-resistant)"  
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;  
Site_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAGG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH_MGC Library."  
220 a 230 c 214 g 171 t  
BASE COUNT 220 a 230 c 214 g 171 t  
ORIGIN  
  
Query Match 20.9%; Score 741.8; DB 10; Length 835;  
Best Local Similarity 98.8%; Pred. No. 1.9e-174;  
Matches 768; Conservative 0; Mismatches 7; Indels 2; Gaps 2;  
  
Qy 1591 gatagctccctgtctagatcattaaagtaacacgaaagagccatctgtgtgtaaaag 1650  
Db 1 GATAGCTCCCTGTCTAGTACATTAAATAACACCAAGAGAGCCATCTGTGTGTAAGAG 60  
  
Qy 1651 gaggccttgccatctctgagagctcttaagctctctcggtctctcgagagacgtgcc 1710  
Db 61 GAGGCTTGGCCATCTGTGAGAGCTCTTTAAGCTCTCGGCTCTCGGAGGACCTGTCC 120  
  
Qy 1711 agtgactcggagagtcattctccagaagagccagctcgcgtctgcgcccagcagccttc 1770  
Db 121 AGTGACTCGGAGAGTCATCTCCAGAGAGAGCCAGCTCCGCTGTGCGCCAGCAGCCTTC 180  
  
Qy 1771 aggaggcgagcaaacacctgtgtaattcccttcgaatgccagggaacctcccaacct 1830  
Db 181 AGGAGGGGAGCAAAACACCTGAGTCACCTCCCATTCGATGCCAGGAACCTCCCAACCT 240  
  
Qy 1831 gcccggggtcccggggttcgaaagaaacattagaggtatcactcaagtagagaca 1890  
Db 241 GCCCGGGGTCCCGGGGTTCGCAAGGAACTTATGAGGTATCACTCAGTAGGACACA 300  
  
Qy 1891 gagacgcctcatgaacgaagaccttgaatccaaagcaaacacctcttggtattctggt 1950  
Db 301 GAGAGCGCTCATGAACGAAGGACTTTGAATCCAAAGCAAAACCATCTGGTATCTGGT 360  
  
Qy 1951 gggactcctgtgaagacccggagagcattctcctggaggcagcagatattctccgagtagcc 2010  
Db 361 GGGACTCTCTGTGAAGACCCGGAGGCAATCTCTGGAGGAGCAGAGATATCTCCGAGTAGCC 420  
  
Qy 2011 accccgcagaagcgctgcgattctccagcagatagaagattattcagagctggagag 2070  
Db 421 ACCCCGCAGAAAGCGCTGCGATCTTCAGCAGAGATATGAAGATTATTCAGAGCTGGGAGAG 480  
  
Qy 2071 ctccccacagatctctttagaaccagttgtgaagatggcctctggccccccacca 2130  
Db 481 CTTCCTCCACCATCTCTTTAGAACACAGTTGTGAAGATGGGCCCTTTGGCCCCCACCACCA 540  
  
Qy 2131 gaggaaagaaaaagacatctcgtgagctcgcagagctgtggtgcaaaaggctattcttcaa 2190  
Db 541 GAGGAAAGAAAAAGGACATCTCTGTGAGCTCCGAGAGCTGTGGCAAAAGGCTATTCTTTCAA 600  
  
Qy 2191 cagatactgctgtagaaatggagaagaaaaatacagaagctccaagcctctgaaatgat 2250  
Db 601 CAGATATCTGCTGTAGAAATGGAGAAGAAAAATCAGAAGCTCCAAAGCTCTGAACATGAT 660  
  
Qy 2251 ttgctgaacagcgcttgaagctcgattatgaagaaataactcctctgtcttaagaagagta 2310  
Db 661 TTGCTGAACAGCGCTTGAAGCTCGATTATGACGAATTAATCTCCTGTCTTAAGAGAGTA 720  
  
Qy 2311 actacagtgt-gggaaagatgctta-gcactccaggaagaatcaaaaattagttg 2365  
Db 721 ACTACAGTGTGGGAACAGATGCTTAGGCATCTCCAGGAAGATCAACAATCATCGGTTGC 777
```

RESULT 6
BG746377
LOCUS
DEFINITION
602703644F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4856891 5',
mRNA sequence.
BG746377
ACCESSION
BG746377
VERSION
BG746377.1 GI:14057030
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 800)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI708 row: h column: 12
High quality sequence stop: 762.
FEATURES
Location/Qualifiers
1..800
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4856891"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 213 a 218 c 203 g 166 t
ORIGIN

Query Match 20.8%; Score 738.8; DB 10; Length 800;
Best Local Similarity 98.6%; Pred. No. 1.1e-173;
Matches 787; Conservative 0; Mismatches 7; Indels 4; Gaps 4;

Qy 1502 agagattcttaacagagctctttagaagaattttgtcccggtgtaataaagccagagcc 1561
Db 2 AGAGATCTTTAAACAGAGCTCTTTAGAAAGATATTTGTCCGGGGTAATAAGCCAGAGGCC 61

Qy 1562 tgacggaacactccatcagtggtgagtagctccctctgtctagtagacattaaagtaaca 1621
Db 62 TGCAGGAACACTCCATCAGTGTGATCTGGATCTGGATCTCCCTGTCTAGTACATTAAAGTAACA 121

Qy 1622 ccagcaagagcattctgtgtgaagaagagccttgccatctctgagagctcccttta 1681
Db 122 CCAGCAAGAGCCATCTGTGTGTGAAGAGGAGGCTTGCCCATCTCTGAGAGCTTCCTTTA 181

Qy 1682 agctctcgcgtctcctcgaggagacgtctccagtgactcggagtagtcatctcccagaagagc 1741
Db 182 AGCTCTCGCTCTCTCGGAGGACCTGTCCAGTGACTCGGAGAGTCATCTCCAGAGAGC 241

Qy 1742 cagctcgcgtctcgcgccccagcagccttcagggagcagaaaacacctgagtcacttc 1801
Db 242 CAGCTCCGCTGTGCGCCCGCAGCAGGCTTCAGGAGCGGCAACACCTCTGAGTCACTTCC 301

Qy	1802	ccatcaaatgccagaaacctccacaacctgcccgggggtccccgggggtttccgaaggga	1861
Db	302	ccatcaaatgccagaaacctccacaacctgcccgggggtccccgggggtttccgaaggga	361
Qy	1862	aacttatgaggtatcaactcagtgagcacagagagcctcatgaacgaagagacttgaat	1921
Db	362	aacttatgaggtatcaactcagtgagcacagagagcctcatgaacgaagagacttgaat	421
Qy	1922	ccaaagcaaacattcttggtgattcttggtgggactctctgtgaaccccgaggcaattcct	1981
Db	422	ccaaagcaaacattcttggtgattcttggtgggactctctgtgaaccccgaggcaattcct	481
Qy	1982	ggaggcagcagatatctcccgagttagcccccgcgagaaggcgtcgattcttcacga	2041
Db	482	ggaggcagcagatatctcccgagttagcccccgcgagaaggcgtcgattcttcacga	541
Qy	2042	g-atatgaagattattcagagctggggaggtttccccacagatctctttagaaccagtt	2100
Db	542	gcataatgaagattattcagagctggggagaggtttccccacagatctctttagaaccagtt	601
Qy	2101	tgtgaag-atggggcctttggccccccaccagagagaaagaaaaggacatctcgtagct	2159
Db	602	tgtgaacgattggccctttggccccccaccagagagaaagaaaaggacatctcgtagct	661
Qy	2160	ccgagagctgtggcaaaaggctattcttcaacagatactgctgcttagaatgga-gaagg	2218
Db	662	ccgagagctgtggcaaaaggctattcttcaacagatactgctgcttagaatgga-gaagg	721
Qy	2219	aaatcagaagctccaagcctcgaatatgatttgctg-aaacagccctgaagctcgat	2277
Db	722	caaatcagaagctccaagcctcgaatatgatttgctg-aaacagccctgaagctcgat	781
Qy	2278	tatgaagaataactctccc	2295
Db	782	catgaagaataactctccc	799

```

RESULT 7
BM476629
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
COMMENT

1121 bp mRNA linear EST 05-FEB-2002
AGENCOURT_6476337 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5553589
5', mRNA sequence.
BM476629 BM476629.1 GI:18525671
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1121)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lou Staudt
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12270 Row: m Column: 14
High quality sequence start: 6
High quality sequence stop: 636.
Location/Qualifiers
1. .1121
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5553589"
/clone_lib="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
FEATURES
source

```

/note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; cloned unidirectionally; oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies Note: this is a NTH MGC Library."

BASE COUNT	312 a	277 c	280 g	251 t	1 others
ORIGIN					

	Query Match	20.7%; Score 737.4; DB 10; Length 1121;
	Best Local Similarity	96.1%; Pred. No. 2.8e-173;
	Matches 799; Conservative	0; Mismatches 26; Indels 6; Gaps 4
Qy	830	acaatattgtcagccacagatatcgaggaaaatcgaaactatgcttccagattggcc 889
Dd	31	ACTATTGTCGCCCCACAGATCGAGGAAAATCGAACTATGTCTCTCAGATGGCC 90
Qy	890	agtctgaagttaacctcatcagtcagtcgacccccaaaaaatagcat-tggagaaaaatttt 948
Dd	91	AGTCTGAAGTTTACCCTCATCAGTCTGCACACCAAAAAAATAGCATNTGGAGAAAAATTTT 150
Qy	949	aaggagatatccttttgctctcaggccatcacagacgtaggccactttggggttttatctgt 1008
Dd	151	AAGGAGATATCCTTTTGCTCTCAGGGCATCAGACAGTGGACCACTTTGGGTTTATCTGT 210
Qy	1009	cgggagcttccggaggtggcggtcttcaatttgtctgttacgtgtttcagtgccaaat 1068
Dd	211	CGGAGTCTTCGGAGTGGCGGCTTTCATTTTGTCTGTACGTGTTTCAGTGCAAAAT 270
Qy	1069	gagcctctggtgatgaaattatgatgaccctgaaacaggccctcacggtggccgcagtg 1128
Dd	271	GAGGCTCTGTTGTATGAAATTATGATGACCTGTAACAGGCGCTCACGGTGCCCCAGTG 330
Qy	1129	cagcagacagctaaggcgccagccagctgctgtgagggctgcccctgc aaagcctgcac 1188
Dd	331	CAGCAGACAGCTAAGGCCCGCCAGCGTGTGTGAGSGCTGCCCTTCGAAAGCCTGCAC 390
Qy	1189	aagctctgtgagaggtagagggaaatgaaattotttccaaacaaactagaactgc aaaag 1248
Dd	391	AAGTCTGTGAGAGGATAGAGGGAATGAATCTTCCAAACAAAACTAGA ACTGCAAAAG 450
Qy	1249	cactgcagcattaaccaactcaggagcagcgactatttttgaagaggttcagaaattg 1308
Dd	451	CACCTGACGACATTAACCAATCAGAGCAGGCGCACTATTTTTGAAGAGGTTTCAGAAATG 510
Qy	1309	agaccgaaatagcagcgagagaatgaattgattattcttttctgagatgtttatat 1368
Dd	511	AGACCGAAATGASCACGAGAGAATCAATTGATTATTTCTTTCTGAGATGTTTATAT 570
Qy	1369	gaagagaacagagaagaacacatccaatttggggagatgaagcagacatcgagatggca 1428
Dd	571	GAAGAGAACAGAAAGAACACATCCATATTTGGGGAGATGAAGCAGACATCGCAGATGGCA 630
Qy	1429	gcagagaataitggagtgaaattaccacccagtcgactcgatttagctagatatgtcg 1488
Dd	631	GCAGAGNAATITGGAAGTGAATTACCCAGTGCCCTCGATTATAGGCTAGATATGCTG 690
Qy	1489	aaaacaagcaaagagatctttaacagagtcttttagaaagtattttgtccgggggtaat 1548
Dd	691	AAAAACAAGCAAAGAGATCTTTAACAGAGTCTTTAGAAAGTATTTTGTCCCGGGGTAA 750
Qy	1549	aaagccagagcctgcaggaacacatcccatacagtg- --ggactctgagatgacct- gtc 1604
Dd	751	AAAGCCAGAGCCCTCAGGGAAACCTCCATCCCCTGTGGGATCTGGTAGCTCCCTGGTC 810
Qy	1605	tagtaca-ttaagtaaacaccagcaagagccatctgtgtgtaaagaggg 1654
Dd	811	TAGTACATTTTAACTAACACCAAGCAAAAGAGCCATCCTGTGTCTGTA AAAAGGG 861

RESULT
BM460573
LOCUS

```

DEFINITION  AGENCOURT_6421241 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5531908
5', mRNA sequence.
ACCESSION   BM460573
VERSION     BM460573.1 GI:18509613
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 1109)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM1214 row: f column: 05
            High quality sequence stop: 630.
FEATURES    source
            1..1109
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:5531908"
            /clone_lib="NIH_MGC_71"
            /tissue_type="leiomyosarcoma"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
            Average insert size 2.1 kb."
BASE COUNT  301 a 311 c 258 g 239 t
ORIGIN
1
|||||
Query Match      20.5%   Score 729.8; DB 10; Length 1109;
Best Local Similarity 97.5%   Pred. NO. 2.2e-171;
Matches 773; Conservative 0; Mismatches 17; Indels 3; Gaps 3;

QY 1340 tgattattctttctgagatgtttatgaagaaacaaagaaacacacatccatttg 1399
|||||
DB 13  tgaattattctttctgagatgtttatgaagaaacaaagaaacacacatccatttg 72
|||||
QY 1400 gggagatgaagcagacatcgagatggcagagagaattatggagtgaaattaccacca 1459
|||||
DB 73  gggagatgaagcagacatcgagatggcagagagaattatggagtgaaattaccacca 132
|||||
QY 1460 gtgcacactcgattaggctgatgatctgtaaaacaaagcagaagagatctttaacagagt 1519
|||||
DB 133 gtgcacactcgattaggctgatgatctgtaaaacaaagcagaagagatctttaacagagt 192
|||||
QY 1520 ctttgaagaattttgtcccggtgtaataaagcagagagcctgcagaaacactccatca 1579
|||||
DB 193 ctttgaagaattttgtcccggtgtaataaagcagagagcctgcagaaacactccatca 252
|||||
QY 1580 gtgtgagatctgtagctccctgtctagatattgaataacacacagcagaagcagctctg 1639
|||||
DB 253 gtgtgagatctgtagctccctgtctagatattgaataacacacagcagaagcagctctg 312
|||||
QY 1640 tgtgtgaagaagagccttgcccatctctgagagctctcttaagctctctcgctctctgg 1699
|||||
DB 313 tgtgtgaagaagagccttgcccatctctgagagctctcttaagctctctcgctctctgg 372
|||||
QY 1700 aggaactgtccagtaactcgagagatctctccagaagcagcagctcgctgtgcgcc 1759
|||||
DB 373 aggaactgtccagtaactcgagagatctctccagaagcagcagctcgctgtgcgcc 432
|||||
QY 1760 agcaggccttcaggagcgagcaaacacccctgagtcacttccccatcgaaatgccaggagac 1819
|||||

```

```

Db 433 AGCAGGCTTCAGGAGCGAGCAAAACACCCCTGAGTCACCTCCCATCGAATGCCAGAAC 492
QY 1820 ctccaaacctgcccgggggtcccccgggggtttccaaagaaacttatgaggtatcaact 1879
|||||
Db 493 CTCCACAAACCTTCGCCGGGGTTCGCCGGGGTTCGCAAGAGAACTTATGAGGTATCACT 552
|||||
QY 1880 cagtgacacagagacgctcatgaacaaagacatttgaatccaaagcaaacacacatttg 1939
|||||
Db 553 CAGTGAGCACAGAGAGCGCTCATGAACGAAGAGACTTTGAATCCAAAGCAAAACCATCTTG 612
|||||
QY 1940 gtgattctgtgtgggactcctgtgaagaccggagggagcattctctggagcagcagatattcc 1999
|||||
Db 613 GTGATTCGTGGGACTCCTGTGAAGACCCCGGAGCATTCCTGGAGCGAGCATATTCC 672
|||||
QY 2000 tccgagtagccaccccgagcagagggcgctgattctctccagcagatataagattattcag 2059
|||||
Db 673 TCCGAGTAGTACCCACCCGAGAGCGGGCGATTCCTCAGCAGATATGAGATTTATCCG 732
|||||
QY 2060 agct-ggagagactccccccagatc-tcctttagaacacagttgt-gaagatgggacct 2116
|||||
Db 733 AGCTGGGAGAGACTTCCCTCCCACTTCCCTTAGAACAGTTGTGGNAGATGGGGCC 792
|||||
QY 2117 ttggccccccacc 2129
|||||
Db 793 TTTGGCGCCTCC 805
|||||

RESULT      9
BE299948    BE299948      600944494F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960782 5',
LOCUS       mRNA sequence.
DEFINITION  BE299948      600944494F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960782 5',
LOCUS       mRNA sequence.
ACCESSION   BE299948
VERSION     BE299948.1 GI:9183696
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 952)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Ling Hong/Rubin Laboratory
            CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LICM54 row: k column: 23
            High quality sequence stop: 709.
FEATURES    Location/Qualifiers
            1..952
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:2960782"
            /clone_lib="NIH_MGC_17"
            /tissue_type="rhabdomyosarcoma"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
            Site_2: XhoI; CDNA made by oligo-dT priming.
            Directionally cloned into EcoRI/XhoI sites using the
            following 5' adaptor: GGCACGAG(G). Size selected by
            for average insert size 1.8kb. Library constructed by
            Ling Hong in the laboratory of Gerald M. Rubin (University
            of California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT  249 a 240 c 250 g 212 t 1 others
ORIGIN
1
|||||

```

Query Match 20.3% Score 722.2; DB 10; Length 952;
Best Local Similarity 97.2%; Pred. No. 1.7e-169;
Matches 766; Conservative 0; Mismatches 19; Indels 3; Gaps 3;

QY 1879 tcagtgcacagacagcctcatgaacgaagcatttgaatccaaagcaaacattctt 1938
Db 1 TCAGTGAGCACAGACAGCGCTCATGAAGCAAGAGGACTTTGAATCCAAAGCAACCATCTT 60

QY 1939 ggtgattctggtgggactcctgtgaagaccggaggcattcctctggaggcagcagatttc 1998
Db 61 GGTGATCTGTGGGACTCCTGTGAAGACCCGGAGGACTTCCCTGGAGGCGACAGATATTC 120

QY 1999 ctccgagtaccaccaccgacgaagcgtgcattctccagcagatatgaagattattca 2058
Db 121 CTCCGAGTAGCCACCCGCGAGAGGGGTGCGATCTTCCAGCAGATATGAGATTTATCA 180

QY 2059 gagctgggagagcttcccccaagatctcctttagaaccagttgtgaaagtggcccttt 2118
Db 181 GAGCTGGGAGAGCTTCCCCCAGCATCTCTTTAGAACCCAGTTGTGAAGATGGCCCTTT 240

QY 2119 ggcctccaccagaggaagaaagacatctcgtgagctcagagcgtgtggcaaaag 2178
Db 241 GGCCCCCACCAGAGAAAGAGACATCTCGTGAGCTCCGAGAGCTGTGGCAAAAG 300

QY 2179 gctattctcaacagatactcgtctagatgagaagaaatcagaagctcccaagcc 2238
Db 301 GCTATCTTCAACAGATAGTCTGCTTAGAATGGAGAGGAAATCAGAAGCTCCAGCC 360

QY 2239 tctgaaatgatttgcctgaacagcgcctgaagctcgattatgaagaaattactcctgt 2298
Db 361 TCTGAAATGATTGCTGAACAAAGCGCTGAAGCTCGATTATGAAGAAATTAATCTCCTGT 420

QY 2299 cttaagaagtaactacagctgtggaaagatccttagcactccaggaagatcaaaatt 2358
Db 421 CTTAAGAAGTAACAGTGTGGGAAAGATGCTTAGCACTCCAGGAAGATCAAAAT 480

QY 2359 aagttgacatgaaataatgcactcgtgtgtggcgaaggtgcccacgtcatcacga 2418
Db 481 AAGTTTGACATGAAATAATGCCTCGGCTG-TGGCAAGGTGTGCCACGTATCACCGA 539

QY 2419 ggtgaaatctggaaattctagctgagcgaattccaccttaaacacagtttcccgacaa 2478
Db 540 GGTGAAATCTGGAATTTCTAGCTGAGCAATCCACCTTAACACCAAGTTTCCAGCAAA 599

QY 2479 cagcagcaagagatgtgccatacaagaactcttaagcagctgcacttcccgacagcat 2538
Db 600 CAGCAGCAAGAGATGTCCATACAAAGAACTCTTAAGCAGCTGACTTCCAGCAGCAT 659

QY 2539 gcgattctta-ttgaccttggggaaaccttccctacacacccactactctctgccagct 2597
Db 660 GCGATTCTTATTGACCTGNGGCGAACCTTTCCTACACACCCATCTCTCTGCCAG-T 718

QY 2598 tggagcagcagcagctatcgctttacaacattttgaagggcctactcacttctagaccagga 2657
Db 719 TGGAGCAGCAGCAGCTATCGCTTTACAAATTTGTGAGGGCTACTCCACTTCTAGACAGGA 778

QY 2658 agtgggat 2665
Db 779 GTGGGTAT 786

RESULT 10
BI907513
LOCUS 838 bp mRNA linear EST 16-OCT-2001
DEFINITION 603065640F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5214945 5',
mRNA sequence.
ACCESSION BI907513
VERSION BI907513.1 GI:16170347
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 838)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>;
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLML1539 row: 0 column: 10
High quality sequence start: 4
High quality sequence stop: 780.
Location/Qualifiers
1. 838
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone IMAGE:5214945"
/clone_lib="NIH_MGC_118"
/tissue_type="leukocyte"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source leukocytes from anonymous pool of non-activated adult donors. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 1.2-3.3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 027. Note: this is a NIH_MGC Library."
BASE COUNT 237 a 199 c 191 g 211 t
ORIGIN
Query Match 20.1%; Score 713.8; DB 10; Length 838;
Best Local Similarity 97.0%; Pred. No. 2e-167;
Matches 812; Conservative 0; Mismatches 17; Indels 8; Gaps 8;
QY 2226 gaagctccaaagcctcgaataatgatttgcgaacagcgcctgaagctcgtattatgaaga 2285
Db 9 GAAGCTCCAAAGCCTCTGAAATGATTGCTGGAACAAGCGCCTGAAGCTCGATTATGAAGA 68
QY 2286 aattactccctctctaaagaagtaactacagctgtgggaaagatgcttagcactccag 2345
Db 69 AATTACTCCCTGCTCTAAAGAAGTAAGTACTAGCTGTGGGAAAGATGCTTAGCAGCTCCAG 128
QY 2346 aagatcaaaaattaaagtttgacatggaataaaatgcactcgcgtgtgggcaaggtgtgac 2405
Db 129 AAGATCAAAAATTAAGTTTGACATGGAAAAATGCACCTCGCTGTGGCAAGGTGTGCC 189
QY 2406 acgtcatcaccgagggtgaatactcggaaattctagctgagcaattccacatttaaacacca 2465
Db 189 ACGTCTATCACCGAGGTGAATCTGGAATTTCTAGCTGAGCAATTCACCTTTAAACACCA 248
QY 2466 gttccacgacaaacagcagcgaagagatgtgccatacaagaactcttaagcagctgac 2525
Db 249 GTTTCCAGCAAAACAGCAGCCAAAGGATGTGCCATACAAAGAACTCTTAAGCAGCTGAC 308
QY 2526 ttccagcagcagcagctattctattgaccttggggcaaccttctactacacacccatactt 2585
Db 309 TTCCAGCAGCATGCGATTCTTA-TGACCTTGGGGAACCTTTCTCTACACACCATACTT 367
QY 2586 ctctgccagcttgggagcagcagcagctatcgctttacaacattttgagcagcctactcact 2645
Db 368 CTCTGCCAGCTTGGAGCAGCAGCAGCTATCGCTTTTACAACA-TTTGAAGGCCCTACTCACT 426
QY 2646 tctagaccaggaagtggtgatattgccaaggtctcagctttgttagcagggcattttgtctct 2705
Db 427 TCTAGACCAAGGAAGTGGGATA-TGCCAAGGCTCTCAGC-TTGTAGCAGGCA-TTGTGCTTCT 483

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: jCSTD/Drp
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLML2063 row: j column: 19
 High quality sequence stop: 811.
 Location/Qualifiers
 1. .832
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5416818"
 /clone_lib="NIH_MGC_87"
 /issue_type="mammary adenocarcinoma, cell line"
 /lab_host="PH10B (phage-resistant)"
 /note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.383 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."
 BASE COUNT 231 a 205 c 186 g 208 t 2 others

FEATURES

source

RESULT 13
 AL570425
 LOCUS
 DEFINITION AL570425 LTI_NFL006.PL2 Homo sapiens cDNA clone CSODI020YF02 5
 prime, mRNA sequence.
 ACCESSION AL570425
 VERSION AL570425.1 GI:12926720
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 851)
 AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source

1. .851
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CSODI020YF02"
 /clone_lib="LTI_NFL006.PL2"
 /issue_type="placenta"
 /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed by
 Life Technologies. Contact : Feng Liang Life Technologies,
 a division of Invitrogen 9800 Medical Center Drive
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : fliang@lifetech.com URL :
<http://fulllength.invitrogen.com>"

BASE COUNT 240 a 214 c 220 g 174 t 3 others
 ORIGIN

Query Match 19.4%; Score 689.8; DB 9; Length 851;
 Best Local Similarity 98.9%; Pred. No. 2e-161;
 Matches 714; Conservative 1; Mismatches 4; Indels 3; Gaps 2;

QY 2836 gattaccacagagacacctctacaat--cacctggagagacagagatcgccacagcctct 2893
 Db 1 GATTACCACAGAGACCTCTCAAAANTCACCTTGGAGGAGACAGAGATCGGCCCGCCCTCT 60
 QY 2894 acgtgccccctgttccctaccatgttgcctcacagttcccgctgggattcgtagcca 2953
 Db 61 ACCTGCCCCCTGGTTCCTCACCATGTTTGCCTCACAGTTCCTCCGCTGGGATTCGTAGCCA 120
 QY 2954 gactcttgatagattttcttcagggaacagaggtcatatttaagtggccttaagtc 3013
 Db 121 GACTCTTGATATGATTTTCTTCAGGGAACAGAGGTCAATATAAGTGGCTTAAGTC 180

QY 2405 cagctcatcaccaggtgaaatctgaaattctagctgagcaattccaccttaaacacc 2464
 Db 29 CAGCTCATCACCAGGTGAAATCTGGAAATTTCTAGCTGAGCAATTCACCTTTAAACACC 88
 QY 2465 agttccccagcaaacagcagccaaagatgtgccatacaaaagactcttaaacagctga 2524
 Db 89 AGTTTCCACGAAACAGCAGCCAAAGGATGTGCCATACAAAGAACTCTTAAAGCAGCTGA 148
 QY 2525 ctccacagcagatcgatctctatgaccttgggggaaaccttctctacacacctact 2584
 Db 149 CTTCCACGACGATGCGATTCTTA-TGACCTTGGCGGAACCTTTCTTACACACCCATACT 207
 QY 2585 tctctgccagctggagcagcagctatcgctttacacatttggagcctactcac 2644
 Db 208 TCTCTGCCAGCTTGGAGCAGGACGATATCGCTTTACACATTTTGAAGCCCTACTCAC 267
 QY 2645 tcttagaccaggaagtggatattgccaaagttctcagctttagcaggcatcttcttc 2704
 Db 268 TTCTAGACCAGGAAGTGGGATATTGCCAAGTCTCAGCTTTGTAGCAGGCATTTGTGCTC 327
 QY 2705 tcatatgagtgagaagagcgttttaaatgtcctaagttctgagtttgacatggggc 2764
 Db 328 TTCAATATGAGTGAGGAAGAGCGCTTTAAATGTCTCAAGTTTCTGATGTTTGACATGGGC 387
 QY 2765 tgcggaacagtatcgccagacatgattttacagatccagatgacagctctcga 2824
 Db 388 TGGGAAACAGATATCGCCAGACATGATTTTACAGATCCAGATGACAGCTCTCGA 447
 QY 2825 ggttgcctcatgattaccacagagacctctacaatcaccttggaggagcagagatggcc 2884
 Db 448 GGTTCCTCATGATTACACAGAGACCTCTACAATCACCTTGGAGGAGCAGAGATCGGCC 507
 QY 2885 ccagctctacgctgccccctgttctcaccatgtttgcctcacagcttcccgctgggat 2944
 Db 508 CCAGCCTCTACGCTGCCCGCTGGTTCCTCACCATGTTTGTGCTCACAGTTCCTCCGCTGGGAT 567
 QY 2945 tcgagcagagcttttgatgattttcttcagggaacagaggtcatatttaa-agt 3003
 Db 568 TCGTAGCCAGAGTCTTTGATATGATTTTCTTCAGGGAACAGAGGTCAATTTAAACAGTG 627
 QY 3004 gctttaagtctgttgggaagccataagcccttgattctgcagcatgaaaaacctagaaacc 3063

```
QY 3014 ttgtggagccataagcccttgattctgcagcatgaaacccctagaaacattgact 3073
DB 181 TGTGGGAGGCATTAAGCCCTTGATCTGCGAGCATGAAGAACTAGAAACCATAGTTGACT 240
QY 3074 ttataaagcacgctaccccaaccttggcttggtacagatggaagacccatcaatcagg 3133
DB 241 TTATAAAAAGCAGCGCTACCCCAACCTTGGCTTGGTACAGATGGAAGAACCATCAATCAGG 300
QY 3134 tatttgaatggacatgctaaacagtttacaagctttatgaagtttgagtaccacgtctctc 3193
DB 301 TATTGAAATGGATCGCTTAACAGATTACAGCTTATGAAGCTTATGAAGTTGAGTACCACGCTCCTTC 360
QY 3194 aagaagaacttatgattctctctctctcagtgacacccaagaatgataaattagaga 3253
DB 361 AAGAAGAATTTATCGATTCTCTCTCTCAGTGACACCAAGAAATGGATTAATTAGAGA 420
QY 3254 aaaccaacagcagcttaccgaaacagaaaccttgacctcttgaacagttgcagtgaggaa 3313
DB 421 AAACCAACAGCAGCTTACGCAAAACAGAACTTGACCTCTTGAACAGTTGCAGGTGGCAA 480
QY 3314 atggtagatcaaaagccttgagcccaattgagaagctcctcctcagcagtgagagcaagc 3373
DB 481 ATGGTAGGATCAAAAGCCTTGAGGCCACCATTTGAGAGCTCCTGAGCAGTGAGAGCAAGC 540
QY 3374 tgaagcagggcattgcttaccttagaaactgagcggtcgccctctgctgcagacggtgagg 3433
DB 541 TGAAGCAGCCATGCTTACCTTAGAACTGAGCGCGGTGGCCCTCTGCTGCAGACGGTGGAGG 600
QY 3434 agctgcggcgagcgagcagccagcgagccggcgagctgagtcgacgcagccgagc 3493
DB 601 AGCTGCGCGCGGA-SGCAGAGCCAGCGAGCGGAGCCTGAGTGACGCGAGCCCGAGC 659
QY 3494 ccagggcgactgacagctctcagggagagattgcaacacacatccacactgtccagggc 3553
DB 660 CCAGCGGCGACTGACGCTCTCAGGAGAGATTGCACACCATCCACACTGTCCAGGCC 719
QY 3554 tt 3555
DB 720 TT 721

RESULT 14
BE300005
LOCUS 60944594F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960783 5',
DEFINITION mRNA sequence.
ACCESSION BE300005
VERSION BE300005.1 GI:9183753
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS NIH-MGC http://mhc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1C054 row: k column: 24
High quality sequence start: 2
High quality sequence stop: 753.
Location/Qualifiers
1. .754
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2960783"
```

```
/clone_lib="NIH_MGC_17"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT	227 a	186 c	177 g	164 t
ORIGIN				
Query Match	19.2%	Score 683.4	DB 10	Length 754
Best Local Similarity	99.1%	Pred. No. 7.5e-160		
Matches	750	Conservative 0	Mismatches 1	Indels 6
Gaps	6			

```
QY 1878 ctcaagtgcagcacagagcgcctcatgaagaaagactttgaaatccaaagcaaacattct 1937
DB 1 CTCAGTGAGCAGACAGAGCGCCTCATGAACGAAGAGACTTTGAATCCAAAGCAAAACCATC- 59
QY 1938 tgggtattctgtgtggactcctgtgaagaccgcggagcattcctctggagcgagcagatatt 1997
DB 60 TGGTGATTCTGTGGGACTCCTGTGAAGACCCGGGAGGCAATTCCTGGAGGCAGCAGATATT 119
QY 1998 cctccgagtagcccccgcagagagcggtgcgattcttccagcagatatgaagattattc 2057
DB 120 CCTCCGAGTAGCCACCCCGCAGAGAGGCGTGGATTCTTCCAGCAGATATGAAGATTATTC 179
QY 2058 agaactgggagagcttccccacagctctcctttagaaccagttgtgaagatgggcccctt 2117
DB 180 AGAGCTGGGAGAGCTTCCCCCAGCATCTCTTTAGAACCCAG-TTGTGAAGATGGGCCCT- 237
QY 2118 tggcccccacagagagaaagaaagacatctcgtgagctccgagcgtgtgagcaaaa 2177
DB 238 TGGCCCCCACCAGAGGAAAGAAAGGACATCTCGTGAGCTCCGAGAGCTGTGSCAAA 297
QY 2178 ggcattcttcaacagatactgctgcttagaattgggaagagaaatcagaagctccaaagc 2237
DB 298 GGCATTCTTCAACAGATACTGCTGCTTAGAATGAGAGAAATCAGAAAGCTCCAAGC 357
QY 2238 ctctgaaaatgatttctgaaacagccctgaagctgagctgattatgaagaaattactccctg 2297
DB 358 CTCTGAAAATGA-TTGCTGAACAAGCCCTGAACTCGATTATGAAGAAATTTACTCCCTG 416
QY 2298 tcttaagaagtaactacagtggtgggaaagatgcttagcactccaggaagatcaaaaaat 2357
DB 417 TCTTAAAGAGTAATACTACAGTGTGGGAAAGATGCTTAGCACTCCAGAGATCAAAAAT 476
QY 2358 taagttagacatggaaaaatgcaactcggctgtgtgggcaaggtgtgccagctcatcaccg 2417
DB 477 TAAGTTTGACATGGAATAAATGCACTCGCTGTGTGGCAAGGTGTGCCACGCTCATACCCG 536
QY 2418 aggtgaaatctggaaattcttagctgagcaattccacttaaacacacagtttccacgcaa 2477
DB 537 AGGTGAAATCTGGAAATTTCTAGCTGAGCAATTCACCTTTAAACACACAGTTTCCACGCA 596
QY 2478 acagcagccaaagatgtgcatacaagaactcttaagcagctgacttccccagcagca 2537
DB 597 ACAGCGCCAAAGGATGTGCCATACAAAGAACTCTTAAAGCAGCTGACTTCCAGCAGCA 656
QY 2538 tgcattcttattgacattggcgaaaccttctctacacacccatctctctgtgccagct 2597
DB 657 TGCATTCTTTA-TGACCTTGGCGAAGCTTTCTTACACACCATACTTCTTGTGCCAGCT 715
QY 2598 tggagcaggacgctatcgctttacaacatttgaag 2634
DB 716 TGGAGCAGGACAGCTATCGC-TTACAACATTTTGAGG 751

RESULT 15
```

BE561889
LOCUS
DEFINITION BE561889 892 bp mRNA linear EST 15-AUG-2000
601346094F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:3678887 5',
mRNA sequence.
ACCESSION BE561889
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 892)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LfCM356 row: 1 column: 24
High quality sequence start: 50
High quality sequence stop: 758.
Location/Qualifiers
1..892
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3678887"
/clone_lib="NIH_MGC_8"
/tissue_type="Burkitt lymphoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lymph; Vector: pOFB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 224 a 240 c 242 g 186 t
ORIGIN

Query Match 19.2%; Score 682.6; DB 10; Length 892;
Best Local Similarity 95.3%; Pred. NO. 1.3e-159;
Matches 790; Conservative 0; Mismatches 29; Indels 10; Gaps 8;

QY 1628 aagagccatctgtgtgaaagaggcccttgcccatctctgagagctctttaagctcc 1687
|||||
DB 63 AAGAGCCATCTGTGTGAAAGAGG-CTTGCCCATCTCTGAGAGCTCTTTAAGCTCC 121
|||||

QY 1688 tcggctctcggaggacctgtccagtgactcgagagtgatctctccagaagaccagctc 1747
|||||
DB 122 TCGGCTCTCGGAGGACCTGCTCCAGTGACTCGGAGAGTCACTCCCAAGAGCCAGCTC 181
|||||

QY 1748 cgctgtcgccagcagcgcccttcaggagcgagcaaacacctgagtcactt-ccccatc 1806
|||||
DB 182 CGCTGTGCCCCAGCAGCCCTTCAGGAGGCGAGCAACACCCCTGAGTCATCTCCCCCATC 241
|||||

QY 1807 gaatgccaggaaacctccacaacctgcccgggggtccccgggggttcgcgaaggaacctt 1866
|||||
DB 242 GAATGCCAGGAACCTCCACAACTGCCCCGGGGTCCCGGGGGTTTCGCAAGGAACCTT 301
|||||

QY 1867 atgaggtatcactcagtgagcacagacgcctcctgaacgaaggaactttgaatccaaa 1926
|||||
DB 302 ATGAGGTATCACTCAGTGAGCAGAGAGCGCTCATGACGAAAGGACTTTGAATCCAAA 361
|||||

QY 1927 gcaaacatatttggtgattctgtgtgacctctctgtgaagaccggaggaattcctggagg 1986
|||||
DB 362 GCAAAATTATCTGGTGATCTGTGTGGACTCCTGTGAAGAGCCCGGAGGCATCTCTGGAGG 421
|||||

Search completed: September 9, 2002, 16:41:06
Job time: 9234 sec

QY 1987 cagcagatatctctccgagtagccacccccgagaagcgctgcgattcttccagcagatat 2046
|||||
DB 422 CAGCAGATATTTCTCCGAGTAGCCACCCCGAGAGGCGTGCATTCTTCCAGCAGATAT 481
|||||

QY 2047 gaagattattcagagctgggagagcttccccccagatctcctcttagaacagatttgtaa 2106
|||||
DB 482 GAAAGATTATTTCAGAGCTGGGAGAGCTTCCCCACGATCTCCTTTAGAACCAAGTTGTGAA 541
|||||

QY 2107 gatgggccccttgccccccaccagaggaagaaagagacatctcgtgagctccgagag 2166
|||||
DB 542 GATGGGCCCTTTGGCCCCCACCAGAGAAACGAAAGAGACATCTCGTAGCTCCGAGAG 601
|||||

QY 2167 ctgtggcaaaagcgtattcttcaacagatactcgtcgttagaattggagaaggaataacag 2226
|||||
DB 602 CTGTGGCAAAAGGCTATTCTTCAACAGATACTGCTGCTTAGAATGGCGAAGGAAATCAG 661
|||||

QY 2227 aagctccaaagcctctgaataatgatttgctgaacaaagcgctgaagctcgattatgaaga 2286
|||||
DB 662 AAGCTCCAAAGCCTCTGAAATATGATTTGCTGAACAAGGCCCTGAAGCTCGATTATGAAGAA 721
|||||

QY 2287 attactccctg-tcttaaaagaagtaactacagtgaggga--aaagatgcttagcactcca 2343
|||||
DB 722 ATTACTCCCTGTGTTCTTAAAGAAAGTAACACAGTGTGGGACCAAGATGGCTTAGCACTCCA 781
|||||

QY 2344 ggaagatcaaaaatttaagtttgacatggaaaaaatgcactcggcgtgttgggcaaggtgtg 2403
|||||
DB 782 GGAAGATC-CCCATTTAGTTTGGCTGGTGACAATGC-CTCGGGTGTGTGGCCAGGGGGG 839
|||||

QY 2404 ccacgtcatcaccgaggtgaaatcttgaaatttctagctgagcaattcc 2452
|||||
DB 840 CC-CGTATCACCAGCGCTGAACACTGGG--AGTTTGTAGCTGGGCAATTC 885
|||||

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2002, 14:27:02 ; Search time 706.85 Seconds
(without alignments)
8634.973 Million cell updates/sec

Title: US-09-762-311-3_COPY_171_3725

Perfect score: 3555
Sequence: 1 atgggaacataacattccac.....tccacactgtccagcctt 3555

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3555	100.0	3983	21	AAZ50906 Human TBC-1 cDNA f
2	3555	100.0	3988	21	AAZ50907 Human TBC-1 cDNA f
3	3026.4	85.1	3326	22	AAI57854 Human polynucleoti
4	2940.2	82.7	3727	21	AAZ77404 Human ORF72959
5	2833.8	79.7	3281	22	AAI59640 Human polynucleoti
6	2510.4	70.6	4039	19	AAV05886 Human Tbc-1 gene.
7	2112	59.4	2362	22	AAH18401 Human cDNA sequenc
8	1000.8	28.2	1422	22	AAZ34787 cDNA encoding nove
9	999.4	28.1	2989	22	AAK9496 Human full-length

10	943.2	26.5	1269	22	AAZ34785 cDNA encoding nove
11	785	22.1	883	22	AAH04450 Human cDNA clone (
12	574.2	16.2	849	22	AAK92006 Human cDNA 5'-end
13	574.2	16.2	849	22	AAK93666 Human cDNA-clone r
14	422.6	11.9	17590	21	AAZ50904 Human TBC-1 partia
15	362	10.2	696	20	AAI15229 Human gene express
16	306.2	8.6	5472	23	ABL02013 Drosophila melanog
17	284.4	8.0	99960	21	AAZ50905 Human TBC-1 partia
18	280	7.9	458	20	AAZ40323 Human secreted pro
19	245.6	6.9	8105	23	ABL02012 Drosophila melanog
20	238.4	6.4	555	22	AAH09958 Human cDNA clone (
21	132.8	3.7	3691	22	AAH58616 Human RECAP polynu
22	132.8	3.7	4827	22	AAI59592 Human polynucleoti
23	132.8	3.7	5294	22	AAI57806 Human polynucleoti
24	128.8	3.6	3585	23	ABL12337 Drosophila melanog
25	123.8	3.5	154	23	AAZ58636 cDNA #1312 encodin
26	119.4	3.4	2626	23	ABL18133 Drosophila melanog
27	119.4	3.4	7829	23	ABL18133 Drosophila melanog
28	108	3.0	2424	23	AAZ71793 DNA encoding novel
29	83.8	2.4	1601	21	AAZ3401 Arabidopsis thalia
30	79.6	2.2	6765	23	ABL12336 Drosophila melanog
31	78.4	2.2	300	20	AAZ13067 Human gene express
32	75.2	2.1	1492	21	AAZ38750 Arabidopsis thalia
33	73.8	2.1	1140	21	AAZ97020 Human secreted pro
34	68.4	1.9	3342	22	AAZ32789 Human secreted pro
35	68	1.9	2492	21	AAZ78199 Human cancer assoc
36	65.2	1.8	1831	22	AAH64964 Human secreted pro
37	65	1.8	1117	22	AAI61302 Human polynucleoti
38	65	1.8	2035	22	AAZ59600 Human cell cycle a
39	65	1.8	2482	22	AAI59516 Human polynucleoti
40	65	1.8	2666	22	AAH17866 Human cDNA sequenc
41	62.6	1.8	2531	23	ABL07359 Drosophila melanog
42	62.6	1.8	4661	23	ABL07358 Drosophila melanog
43	61.6	1.7	2801	22	AAZ32643 Human cDNA encodin
44	61.6	1.7	2803	22	AAZ27080 cDNA encoding nove
45	61.6	1.7	2903	22	AAZ51641 Human polynucleoti

ALIGNMENTS

RESULT 1

AAZ50906
ID AAZ50906 standard; cDNA; 3983 BP.

XX AAZ50906;

XX 31-MAY-2000 (first entry)

XX Human TBC-1 cDNA from first transcript.

XX TBC-1; human; biallelic marker; chromosome 4; cell cycle regulator; SNP;
XX Single nucleotide polymorphism; tissue differentiation; prostate cancer;
XX linkage analysis; genetic map; detection; diagnosis; genotyping;
XX transgenic animal; screening; alternative splicing; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT 5'UTR 1..170

FT /tag= a

FT /note= "Spans through exon 1 and part of exon 2"

FT CDS 171..3677

FT /tag= b

FT /product= "Human TBC-1 protein"

FT /note= "First transcription product"

FT 3'UTR 3726..3983

FT /tag= c

FT polyA_signal 3942..3947

FT /tag= d

XX WO200008209-A2.

XX

KW Single nucleotide polymorphism; tissue differentiation; prostate cancer;
 KW linkage analysis; genetic map; detection; diagnosis; genotyping;
 KW transgenic animal; screening; alternative splicing; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT 5'UTR 1..175

FT /*tag- a

FT misc_feature 1..438

FT /*tag- b

FT /*note- "Homologous to 5' EST obtained from human

FT pancreas cDNA library"

FT CDS 176..3682

FT /*tag- c

FT /*product- "Human TBC-1 protein"

FT /*note- "Second transcription product"

FT 3'UTR 3731..3988

FT /*tag- d

FT polyA_signal 3947..3952

FT /*tag- e

XX WO200008209-A2.

PN 17-FEB-2000.

XX 06-AUG-1999;

XX 99WO-IB01444.

XX 07-AUG-1998;

XX 98US-0095653.

XX (GEST) GENSET.

PA Blumenfeld M, Bougueret L, Chumakov I;

PI WPI; 2000-205736/18.

XX P-PSDB; AAY45096.

DR New isolated human TBC-1 nucleic acids, useful for developing products

XX for the diagnosis and treatment of disorders involving cell

XX proliferation, particularly prostate cancer

XX Claim 4; Page 156-161; 166pp; English.

XX The present sequence is the cDNA encoding the second transcription

CC product of human TBC-1 protein, comprising exons 1bis to 2 and A-L.

CC Alternative splicing events result in two mRNA molecules, due to

CC splicing at two distinct first exons, exon1 and exon 1bis. TBC-1 gene

CC is mapped to a candidate region of prostate cancer on chromosome 4.

CC TBC-1 gene is involved in the regulation of cell cycle and tissue

CC differentiation in mammals. An alteration of TBC-1 sequence may be

CC associated with a pathological condition, resulting in abnormal cell

CC proliferation leading to cancer, e.g. prostate cancer. The biallelic

CC markers can be used for generation of genetic maps, linkage analysis and

CC association studies. TBC-1 sequence can be used for detection,

CC diagnosis, genotyping, production of transgenic animals and screening

CC of compounds for use in therapy.

XX Sequence 3988 BP; 1107 A; 983 C; 1002 G; 896 T; 0 other;

Query Match

Best Local Similarity 100.0%

Matches 3555; Conservative 0;

Mismatches 0; Indels 0; Gaps 0;

Length 3988;

Pred. No. 0;

1 atggaaccataacattcacagaagaacattcttcttaacagaggtctcgttgat 60

176 atggaaccataacattcacagaagaacattcttcttaacagaggtctcgttgat 235

61 ttggcctcagctggtgggtccctcgtgctgcttcctaccacacatgccatgctg 120

236 ttggcctcagctggtgggtccctcgtgctgcttcctaccacacatgccatgctg 295

121 ccttggttggtaggtgaggaactcagcagcagcagtcaccacagaagaacctgta 180

296 ccttggttggtaggtgaggaactcagcagcagcagtcaccacagaagaacctgta 355
 181 accaagcaagtcggttgcgtttcaccctctggtgactgagatgtgaacctgagccagg 240
 356 accaagcaagtcggttgcgtttcaccctctggtgactgagatgtgaacctgagccagg 415
 241 agaagtcacagtcggatccctcgtatcttccagcatctttgagtcgaagcctcagct 300
 416 agaagtcacagtcggatccctcgtatcttccagcatctttgagtcgaagcctcagct 475
 301 gttcacaaactgattcacacagtcaccccaagttactttctgtctgattaaggaa 360
 476 gttcacaaactgattcacacagtcaccccaagttactttctgtctgattaaggaa 535
 361 gacgtgtccacggcagagtcattctgttctcaaaagccatgatacaaaagaagtg 420
 536 gacgtgtccacggcagagtcattctgttctcaaaagccatgatacaaaagaagtg 595
 421 cctgagatcatcagtcctcctcgttcagggggaagatcccgccagcagagagctgcac 480
 596 cctgagatcatcagtcctcctcgttcagggggaagatcccgccagcagagagctgcac 655
 481 tgccgtccgagtcgacacacgcttttccaaagaattcagagtgctcttctcggccgc 540
 656 tgccgtccgagtcgacacacgcttttccaaagaattcagagtgctcttctcggccgc 715
 541 gtgacgtgtgcccacaaagagtcgcggccctcgtatcgcagagtcgacagagagttc 600
 716 gtgacgtgtgcccacaaagagtcgcggccctcgtatcgcagagtcgacagagagttc 775
 601 aatcacgtcagcggcagccggggtccgagagcccccgcaccccccacccatgcgcg 660
 776 aatcacgtcagcggcagccggggtccgagagcccccgcaccccccacccatgcgcg 835
 661 ccacagggcagccagagcctgtgcgagcccatgcgaagtccttctccacagccggc 720
 836 ccacagggcagccagagcctgtgcgagcccatgcgaagtccttctccacagccggc 895
 721 ctgctcgtcgtccttttaggaagagtcgaggtcggggccctccgagcagcagccttc 780
 896 ctgctcgtcgtccttttaggaagagtcgaggtcggggccctccgagcagcagccttc 955
 781 ttacgtccttcgagagagcagcattgagaccactcattagcggacacataatgtg 840
 956 ttacgtccttcgagagagcagcattgagaccactcattagcggacacataatgtg 1015
 841 cagccacagatcagaggaataatgaaatcgtcttccacgattgcccagtcgaagtt 900
 1016 cagccacagatcagaggaataatgaaatcgtcttccacgattgcccagtcgaagtt 1075
 901 tacctcagtcctgacaccccaaaaaatagattggagaaaaattttaagagatctcc 960
 1076 tacctcagtcctgacaccccaaaaaatagattggagaaaaattttaagagatctcc 1135
 961 ttttctcagggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1020
 1136 ttttctcagggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1195
 1021 ggagtcggcggttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1080
 1196 ggagtcggcggttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1255
 1081 gatgaattatgatgacctgaaacagggcttcagcgttgcccgagtcgacagacagct 1140
 1256 gatgaattatgatgacctgaaacagggcttcagcgttgcccgagtcgacagacagct 1315
 1141 aagcgcagccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1200
 1316 aagcgcagccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1375
 1201 aggatagaggggaatgaattcttccaaaaacaaactagaactgcaaaagcaccatgcac 1260

||||| 943 cagatggcagcagagaattatggaattaccaccagtcgaactgatttaggcta 1002
QY 1480 gatctgctgaaaaaacaagcaagagatctttaacagagctctttaaaagtattttgtcc 1539
Db 1003 gatctgctgaaaaaacaagcaagagatctttaacagagctctttaaaagtattttgtcc 1062
QY 1540 cggggttaataaaagccagagcctgaggaacacctcatcagtgatctgatatgactcc 1599
Db 1063 cggggttaataaaagccagagcctgaggaacacctcatcagtgatctgatatgactcc 1122
QY 1600 ctgtctagtacattaagtaaacacagcaagagccatctgtgtgtaaaagggagccttg 1659
Db 1123 ctgtctagtacattaagtaaacacagcaagagccatctgtgtgtaaaagggagccttg 1182
QY 1660 ccatctctgagagctcctttaagctcctggtcctcctgagagagatgtctcagtgactcg 1719
Db 1183 ccatctctgagagctcctttaagctcctcctggtcctcctgagagagatgtctcagtgactcg 1242
QY 1720 gagagtcattcccaagaagcagatccgctgtcgcaccagcagcagccttcaggagggcga 1779
Db 1243 gagagtcattcccaagaagcagatccgctgtcgcaccagcagcagccttcaggagggcga 1302
QY 1780 gcaaacacccctgagtcacttcccatcgaatgccaggaacctcccaacctgcccggggg 1839
Db 1303 gcaaacacccctgagtcacttcccatcgaatgccaggaacctcccaacctgcccggggg 1362
QY 1840 tccccgggggttcgaaagaaacttatgaggtatcaactcagtgagcagagagagcct 1899
Db 1363 tccccgggggttcgaaagaaacttatgaggtatcaactcagtgagcagagagagcct 1422
QY 1900 catgaacgaagagactttgaatccaaagcaaacctcttgggtgattctggtggagcctcct 1959
Db 1423 catgaacgaagagactttgaatccaaagcaaacctcttgggtgattctggtggagcctcct 1482
QY 1960 gtgaagaccggaggcattcctggagcagcagatattcctcagtagcagcccccgcag 2019
Db 1483 gtgaagaccggaggcattcctggagcagcagatattcctcagtagcagcccccgcag 1542
QY 2020 aagcgtgcgattctccagcagatgatgaagattatcagagctggagagcttccccca 2079
Db 1543 aagcgtgcgattctccagcagatgatgaagattatcagagctggagagcttccccca 1602
QY 2080 cgatctcctttagaacagtttgtgaagatgggcccctttggccccaccagagggaaaag 2139
Db 1603 cgatctcctttagaacagtttgtgaagatgggcccctttggccccaccagagggaaaag 1662
QY 2140 aaaagacatctcgtgagctccgagagctgtggcaaaaggctatttctcaacagatactg 2199
Db 1663 aaaaggacatctcgtgagctccgagagctgtggcaaaaggctatttctcaacagatactg 1722
QY 2200 ctgcttagaatggagaaggaatacgaagctccaagcctctgaaaatgatttgcgtgaac 2259
Db 1723 ctgcttagaatggagaaggaatacgaagctccaagcctctgaaaatgatttgcgtgaac 1782
QY 2260 aagcgcctgaagctcgatgatgagaaattactccctgtctttaaagaagtaactacagtg 2319
Db 1783 aagcgcctgaagctcgatgatgagaaattactccctgtctttaaagaagtaactacagtg 1842
QY 2320 tgggaaaagatgtctgagcactccagagaatcaaaaattaaagtttgacatgggaaaaatg 2379
Db 1843 tgggaaaagatgtctgagcactccagagaatcaaaaattaaagtttgacatgggaaaaatg 1902
QY 2380 cactcggctgttgggcaaggtgtgcacagtcatacccgaggtgaaaatcttgaaaatttcta 2439
Db 1903 cactcggctgttgggcaaggtgtgcacagtcatacccgaggtgaaaatcttgaaaatttcta 1962
QY 2440 gctgagcaattccaccttaaacaccagtttcccagcaaacagcagcgaagggatgtgcca 2499
Db 1963 gctgagcaattccaccttaaacaccagtttcccagcaaacagcagcgaagggatgtgcca 2022
QY 2500 tacaagaaactcttaagcagctgaactccagcagcagatcgcgattcttattgacctggg 2559
|||||

RESULT 4
AAC77404

Db 2023 tacaagaactcttaaaagcagctgaacttcccagcagcagctgattcttattgacctggg 2082
QY 2560 cgaactcttctacacacccatacttctctgcccagcttgagcagagacagctatgcctt 2619
Db 2083 cgaactcttctacacacccatacttctctgcccagcttgagcagagacagctatgcctt 2142
QY 2620 tacaacattttgaagccttactcactcttagaccaggaagtggtatgtgccaaaggtctc 2679
Db 2143 tacaacattttgaagccttactcactcttagaccaggaagtggtatgtgccaaaggtctc 2202
QY 2680 agctttgtagcaggcattttgcttcttcatactagtgaggaagagcgttttaaaatgctc 2739
Db 2203 agctttgtagcaggcattttgcttcttcatactagtgaggaagagcgttttaaaatgctc 2262
QY 2740 aagtttctcagtttgacatgggctgcgaaacagatctgcgcacacatgattatttta 2799
Db 2263 aagtttctcagtttgacatgggctgcgaaacagatctgcgcacacatgattatttta 2322
QY 2800 cagatccagatgtaccagctctcaggttgccttctcattgattaccacagagacctacaat 2859
Db 2323 cagatccagatgtaccagctctcaggttgccttctcattgattaccacagagacctacaat 2382
QY 2860 caactggaggagcacagatggcccagcctctacgctgcctcctggcttccctcaacatg 2919
Db 2383 caactggaggagcacagatggcccagcctctacgctgcctcctggcttccctcaacatg 2442
QY 2920 tttgcttcagtttcccgctgggattcgttagcagagctcttctgattgtttttcttcag 2979
Db 2443 tttgcttcagtttcccgctgggattcgttagcagagctcttctgattgtttttcttcag 2502
QY 2980 ggaacagagctcatatttaaagtggtcttaagttctgttggaagcctaagccttgatt 3039
Db 2503 ggaacagagctcatatttaaagtggtcttaagttctgttggaagcctaagccttgatt 2562
QY 3040 ctgagcatgaaacacctagaacacatagttgactttataaaagcagcgtcaccaacctt 3099
Db 2563 ctgagcatgaaacacctagaacacatagttgactttataaaagcagcgtcaccaacctt 2622
QY 3100 ggcctgtgacagatgaaagaccatcaatcaggttatttgaatggacatcgtaaacag 3159
Db 2623 ggcctgtgacagatgaaagaccatcaatcaggttatttgaatggacatcgtaaacag 2682
QY 3160 ttacaagcttatgaagtgtgagtcaccagctcttcaagaagaacttatcgattctctcct 3219
Db 2683 ttacaagcttatgaagtgtgagtcaccagctcttcaagaagaacttatcgattctctcct 2742
QY 3220 ctcaagtacaacaaagaatggataatttagaagaacaaacagcagcttacgcaaacag 3279
Db 2743 ctcaagtacaacaaagaatggataatttagaagaacaaacagcagcttacgcaaacag 2802
QY 3280 aaccttgacctcttgaaacagttgcaggtggcaaatggtaggattccaaagccttgaggcc 3339
Db 2803 aaccttgacctcttgaaacagttgcaggtggcaaatggtaggattccaaagccttgaggcc 2862
QY 3340 accttgagaagctcctgagcagctgagagcaagctgaagcagccatgcttaaccttagaa 3399
Db 2863 accttgagaagctcctgagcagctgagagcaagctgaagcagccatgcttaaccttagaa 2922
QY 3400 ctggagcgttcggccctgctgcagacggttgagagagctgcggcgaggagcagagccc 3459
Db 2923 ctggagcgttcggccctgctgcagacggttgagagagctgcggcgaggagcagagccc 2982
QY 3460 agcgaccggagcctgagtgacgcagcccgagccacagcgagctgacagctctgcagg 3519
Db 2983 agcgaccggagcctgagtgacgcagcccgagccacagcgagctgacagctctgcagg 3042
QY 3520 agagattgcaacacccatcccaactgtccaggcctt 3555
Db 3043 agagattgcaacacccatcccaactgtccaggcctt 3078

ID	AAC77404 standard; cDNA: 3727 BP.	Query Match	82.7%; Score 2940.2; DB 21; Length 3727;
XX		Best Local Similarity	99.3%; pred. No. 0;
AC	AAC77404;	Matches 3005; Conservative	0; Mismatches 13; Indels 7; Gaps 5;
XX			
DT	08-FEB-2001 (first entry)		
XX			
DE	Human ORFX ORF2959 polynucleotide sequence SEQ ID NO:5917.		
XX			
KW	Human: open reading frame; ORFX: detection; cytototoxic; hepatotropic;		
KW	vulnary; antipapillary; antipapillary; antipapillary; antipapillary;		
KW	anticonvulsant; osteopathic; antipapillary; antipapillary; antipapillary;		
KW	immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;		
KW	hypotensive; dermatological; immunosuppressive; antidiabetic;		
KW	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;		
KW	antianaemic; gene therapy; cancer; proliferative disorder; hypertension;		
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;		
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;		
KW	cholesterol ester storage; systemic lupus erythematosus; infection;		
KW	severe combined immunodeficiency; malaria; autoimmune disorder;		
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;		
KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;		
KW	thrombosis; contraceptive; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200058473-A2.		
XX			
PD	05-OCT-2000.		
XX			
XX	31-MAR-2000; 2000WO-US08621.		
PF			
XX	31-MAR-1999; 99US-0127607.		
PR	02-APR-1999; 99US-0127636.		
PR	05-APR-1999; 99US-0127728.		
PR	30-MAR-2000; 2000US-0540763.		
XX			
PA	(CURA-) CURAGEN CORP.		
XX			
PI	Shinkets RA, Leach M;		
XX			
DR	WPI: 2000-602362/57.		
DR	P-PSDB; ABA3195.		
XX			
PT	Novel nucleic acids and peptides derived from open reading frame X,		
PT	useful for treating e.g. cancers, proliferative disorders,		
PT	neurodegenerative disorders and cardiovascular disease -		
XX			
PS	Claim 5; Page 5094-5096; 5507pp; English.		
XX			
CC	AAC74446 to AAC77606 encode the proteins given in ABA40237 to ABA43397,		
CC	which represent the human ORFX open reading frames 1 to 3161. The ORFX		
CC	sequences have activities such as: cytototoxic; hepatotropic; vulnary;		
CC	antipapillary; antipapillary; antipapillary; antipapillary;		
CC	osteopathic; anticonvulsant; antidiabetic; immunosuppressive;		
CC	immunostimulant; cardiac; thrombolytic; coagulant; vasotrophic;		
CC	antidiabetic; hypotensive; dermatological; immunosuppressive;		
CC	antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;		
CC	antithyroid; and antianaemic. The sequences can be used for determining		
CC	the presence of or predisposition to, or preventing or treating		
CC	pathological conditions associated with an ORFX-associated disorder. The		
CC	nucleic acids can be used to express ORFX proteins in gene therapy		
CC	vectors. The proteins and nucleic acids may be used to treat cancers,		
CC	proliferative disorders, neurodegenerative disorders, osteoarthritis,		
CC	graft vs host disease, cardiovascular disease, diabetes mellitus,		
CC	hypertension, hypothyroidism, cholesterol ester storage, systemic lupus		
CC	erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,		
CC	bacterial or fungal infection, malaria, autoimmune disorders, asthma,		
CC	allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,		
CC	nocturnal haemoglobinuria, antiinflammatory disease; to enhance		
CC	coagulation; to inhibit thrombosis; and as a contraceptive.		
XX			
SQ	Sequence 3727 BP; 1031 A; 889 C; 931 G; 875 T; 1 other;		

Oy 1552 gccagagcctgcaggaaacactccatcagttggtgatctggatagctccctgtctagtaca 1611
Db 1024 gccagagcctgcaggaaacactccatcagttggtgatctggatagctccctgtctagtaca 1083
Oy 1612 ttaagtaacacacagcaaaagagccatctgtgtgtgaaagaggccttgcccatctctgag 1671
Db 1084 ttaagtaacacacagcaaaagagccatctgtgtgtgaaagaggccttgcccatctctgag 1143
Oy 1672 agtccctttaagctccctcggctcctcaggagacactgtccagtgactcggagagtcactctc 1731
Db 1144 agtccctttaagctccctcggctcctcaggagacactgtccagtgactcggagagtcactctc 1203
Oy 1732 ccagaagagccagctccgctgtgccccagcagcagccttccaggagcgcagcaaacacccctg 1791
Db 1204 ccagaagagccagctccgctgtgccccagcagcagccttccaggagcgcagcaaacacccctg 1263
Oy 1792 agtcaacttcccactcgatgccaggaacactccacaacctgcgccgggggtcccgggggtt 1851
Db 1364 agtcaacttcccactcgatgccaggaacactccacaacctgcgccgggggtcccgggggtt 1323
Oy 1852 tcgcaagggaaacttatgaggtatcaactcagtgagcacagagacgacctcatgaacgaag 1911
Db 1324 tcgcaagggaaacttatgaggtatcaactcagtgagcacagagacgacctcatgaacgaag 1383
Oy 1912 gactttgaatccaaagcaaacactctgtgtgattctgtgtggactcctgtgaagaccctgg 1971
Db 1384 gactttgaatccaaagcaaacactctgtgtgattctgtgtggactcctgtgaagaccctgg 1443
Oy 1972 aggcattctctggagcgagcagatattctcgcagtagccaccccgagagggcgctgcgat 2031
Db 1444 aggcattctctggagcgagcagatattctcgcagtagccaccccgagagggcgctgcgat 1503
Oy 2032 tcttccagcagatataagattatccagagctggagagcttcccacagatctcccttta 2091
Db 1504 tcttccagcagatataagattatccagagctggagagcttcccacagatctcccttta 1563
Oy 2092 gaaccagttgtgaagtggcctcttggccccccacagaggaagaaagacacatct 2151
Db 1564 gaaccagttgtgaagtggcctcttggccccccacagaggaagaaagacacatct 1622
Oy 2152 cgtgagctccgagagctgtggcaaaaggctattcttcaacagatactgc-tgcttagaat 2210
Db 1623 cgtgagctccgagagctgtggcaaaaggctattcttcaacagatactgc-tgcttagaat 1682
Oy 2211 ggagaaagaaaatcagagctccagcctctgaaaaatgatttgcagaaagcgcctgaa 2270
Db 1683 ggagaaagaaaatcagagctccagcctctgaaaaatgatttgcagaaagcgcctgaa 1742
Oy 2271 gctcgattatgaagaaattactccctgtcttaagaagtaactacagtggtgggaaagat 2330
Db 1743 gctcgattatgaagaaattactccctgtcttaagaagtaactacagtggtgggaaagat 1802
Oy 2331 gcttagcactccagagatcaaaaaataagtttgacatggaaaaaatgcactcggctgt 2390
Db 1803 gcttagcactccagagatcaaaaaataagtttgacatggaaaaaatgcactcggctgt 1862
Oy 2391 tgggcaaggtgtccacgtcatccaggtgaaatctgaaattcttagctagacaatt 2450
Db 1863 tgggcaaggtgtccacgtcatccaggtgaaatctgaaattcttagctagacaatt 1922
Oy 2451 ccactttaaacacacagtttccagcaaacagcagcagcagatgtgcatacaaaagact 2510
Db 1923 ccactttaaacacacagtttccagcaaacagcagcagcagatgtgcatacaaaagact 1982
Oy 2511 cttaagcagctgacttcccagcagcatgogattcttattgaccttggcgaaacctttcc 2570
Db 1983 cttaagcagctgacttcccagcagcatgogattcttattgaccttggcgaaacctttcc 2042
Oy 2571 tacacacccatacttctcgcagcagcttgagagcagcagctatcgtttacaacatttt 2630
Db 2043 tacacacccatacttctcgcagcagcttgagagcagcagctatcgtttacaacatttt 2102
Oy 2631 gaagccttactcacttctagaccaggaagtggatattgccaaggtctcagctttgtagc 2690

Db 2103 gaagccctactcacttcttagaccaggaagtgtggtattgccaaggtctcagctttgtagc 2162
Oy 2691 aggcattttgotttcttcataatgagtgagggaagcgtttaaaaatgctcaagtttctgat 2750
Db 2163 aggcattttgotttcttcataatgagtgagggaagcgtttaaaaatgctcaagtttctgat 2222
Oy 2751 gtttgaacatggggtcgtcggaacacagtatcggccaagacatgattattttacagatccagat 2810
Db 2223 gtttgaacatggggtcgtcggaacacagtatcggccaagacatgattattttacagatccagat 2282
Oy 2811 gtaccagctctcagaggtgtcttcattgattaccacagagaccttacaatccacctggagga 2870
Db 2283 gtaccagctctcagaggtgtcttcattgattaccacagagaccttacaatccacctggagga 2342
Oy 2871 gcacgagatcggccccgcctctctacgtcgtccctcgtgttctctaccatgtttgtcctca 2930
Db 2343 gcacgagatcggccccgcctctctacgtcgtccctcgtgttctctaccatgtttgtcctca 2402
Oy 2931 gttcccgctgggtattcgttagccagagctctttgatatttttctcagggaaacagaggt 2990
Db 2403 gttcccgctgggtattcgttagccagagctctttgatatttttctcagggaaacagaggt 2462
Oy 2991 catattaaagtggtctttaagctctgttgggaagccataaagcccttgattctgcagcatga 3050
Db 2463 catattaaagtggtctttaagctctgttgggaagccataaagcccttgattctgcagcatga 2522
Oy 3051 aaacctagaacacatagttgactttataaaaaagcacgtactccccacaccttggcttggatca 3110
Db 2523 aaacctagaacacatagttgactttataaaaaagcacgtactccccacaccttggcttggatca 2582
Oy 3111 gatggaagaacacatcaatcaggtatttgaattggaatggacatcgtaaacagttacaagctta 3170
Db 2583 gatggaagaacacatcaatcaggtatttgaattggaatggacatcgtaaacagttacaagctta 2642
Oy 3171 tgaagttgagtlaccacgtctctcaagaagaacttatcgattctctctctcagtgacaa 3230
Db 2643 tgaagttgagtlaccacgtctctcaagaagaacttatcgattctctctcagtgacaa 2702
Oy 3231 ccaaaagtggataataatagagaaaaaacacagcagcttaacgcaaacagaccttgacct 3290
Db 2703 ccaaaagtggataataatagagaaaaaacacagcagcttaacgcaaacagaccttgacct 2762
Oy 3291 ccttgaacagttcagctggtgcaaatgttagatccaaagccttgagccaccattgagaa 3350
Db 2763 ccttgaacagttcagctggtgcaaatgttagatccaaagccttgagccaccattgagaa 2822
Oy 3351 gctcctgagcagtgagagcaagctgaagcagggccatgcttaccttagaactggagcggctc 3410
Db 2823 gctcctgagcagtgagagcaagctgaagcagggccatgcttaccttagaactggagcggctc 2882
Oy 3411 ggcctctcgcagacggttgaggagctcgcggcgaggcagcagagccagcagccggga 3470
Db 2883 g--cctgctcagacggttgaggagctcgcggcgaggcagcagagccagcagccggga 2940
Oy 3471 gctcaggtgcacgcagcccgagcccgcccgccgactgcagctcgcagggagagattgcaa 3530
Db 2941 gctcaggtgcacgcagcccgagcccgcccgcccgactgcagctcgcagggagagattgcaa 2999
Oy 3531 caccatcccaactgtccaggcctt 3555
Db 3000 caccat--cacactgtccaggcctt 3022

RESULT 5

AAI59640

ID AAI59640 standard; cDNA; 3281 BP.

XX AAI59640;

AC AAI59640;

XX 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 3629.

XX	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia; ss.
XX	Homo sapiens.
XX	WO200153312-A1.
XX	26-JUL-2001.
XX	26-DEC-2000; 2000WO-US34263.
XX	21-JAN-2000; 2000US-0488725.
XX	25-APR-2000; 2000US-0552317.
XX	09-JUL-2000; 2000US-0598042.
XX	19-JUL-2000; 2000US-0620312.
XX	03-AUG-2000; 2000US-0653450.
XX	14-SEP-2000; 2000US-0662191.
XX	19-OCT-2000; 2000US-0693036.
XX	29-NOV-2000; 2000US-0727344.
XX	(HYSE-) HYSEQ INC.
XX	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D; Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J; Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX	WPI: 2001-442253/47.
XX	P-PSDB: AAM40484.
XX	Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -
XX	Claim 1: SEQ ID NO 3629; 10078pp; English.
XX	The invention relates to human nucleic acids (AAI57798-AAI61369) and the encoded polypeptides (AAM38642-AAM42213) with nootropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.
XX	Note: The sequence data for this patent did not form part of the printed specification.
XX	Sequence 3281 BP; 912 A; 812 C; 824 G; 733 T; 0 other;
QY	Query Match 79.7%; Score 2833.8; DB 22; Length 3281; Best Local Similarity 98.3%; Pred. No. 0; Matches 2983; Conservative 0; Mismatches 32; Indels 20; Gaps 11
QY	541 gtacagtggtgcacacaagaagctccgcggccctgacgcagtgatcgagaagttc 600
Db	1 gtacagtggtgcacacaagaagctccgcggccctgacgcagtgatcgagaagttc 60
QY	601 aatcagtcagcgcgagcggggttcagagagcccccgcacacccgcgcctatgcgcg 660
Db	61 aatcagtcagcgcgagcggggttcagagagcccccgcacacccgcgcctatgcgcg 120
QY	661 cccacagggagccagagagcctgtgcgagggccatgcgcaagtctcttccagccggc 720

QY 1782 aaacacctgagtcacttcccatcgaatgcaggaaacctccacaacctgcgccgggggtc 1841
|||||
Db 1261 aaacacctgagtcacttcccatcgaatgcaggaaacctccacaacctgcgccgggggtc 1320
|||||
QY 1842 cccgggggttcgaaagaaacttatgaggtatcaactcagtgagcacagagagcctcca 1901
|||||
Db 1321 cccgggggttcgaaagaaacttatgaggtatcaactcagtgagcacagagagcctcca 1380
|||||
QY 1902 tgaacgaaggaactttgaatccaaagcaaacatcttgggtgattctgtgggaactcctgt 1961
|||||
Db 1381 tgaacgaaggaactttgaatccaaagcaaacatcttgggtgattctgtgggaactcctgt 1440
|||||
QY 1962 gaaacccggaggaattcctgagggcagacagatatctcctcagtagtagccaccgcagaa 2021
|||||
Db 1441 gaaacccggaggaattcctgagggcagacagatatctcctcagtagtagccaccgcagaa 1500
|||||
QY 2022 ggcgtgcgattcttcacagcagatatgaagattattcagagctggagagcttcccaccag 2081
|||||
Db 1501 ggcgtgcgattcttcacagcagatatgaagattattcagagctggagagcttcccaccag 1560
|||||
QY 2082 atctccttgaacacagttgtgaagtggccctttggcccccaccacagaggaagaa 2141
|||||
Db 1561 atctccttgaacacagttgtgaagtggccctttggcccccaccacagaggaagaa 1620
|||||
QY 2142 aagacatctcgtgagctccagagctgtgcaaaagctattctcaacagatactgt 2201
|||||
Db 1621 aaggaca tctcgtgagctccagagctgtgcaaaagctattctcaacagatactgt 1680
|||||
QY 2202 gcttagaatgagaagaaatcagaagctccaagctcctgaaatgatttgcgtaacaa 2261
|||||
Db 1681 gcttagaatgagaagaaatcagaagctccaagctcctgaaatgatttgcgtaacaa 1740
|||||
QY 2262 gcgcctgaagctgattatgaagaaattactcctctttaaagaagttaactacagtggtg 2321
|||||
Db 1741 gcgcctgaagctgattatgaagaaattactcctctttaaagaagttaactacagtggtg 1800
|||||
QY 2322 ggaagagctctagcctccagaaagatcaaaatgaatttcaatggaagaaatgca 2381
|||||
Db 1801 ggaagagctctagcctccagaaagatcaaaatgaatttcaatggaagaaatgca 1860
|||||
QY 2382 ctcggtcttgggaaggtgtg - ccacgtcatcccgaggtgaaatctggaaatttcttag 2440
|||||
Db 1861 ctcggtcttgggaaggtgtgcccacgtcatcccgaggtgaaatctggaaatttcttag 1920
|||||
QY 2441 ctgagcaattccacttaaacacagtttccacgaaacacagcagccaaagagtgccat 2500
|||||
Db 1921 ctgagcaattccacttaaacacagtttccacgaaacacagcagccaaagagtgccat 1980
|||||
QY 2501 acaagaactcttaagcagctgacttcccagcagcagctgattcttattgaccttgggc 2560
|||||
Db 1981 acaagaactcttaagcagctgacttcccagcagcagctgattcttattgaccttgggc 2040
|||||
QY 2561 gaaccttctacacacccacttctctgcccagcttggagcaggacagctatcgcttt 2620
|||||
Db 2041 gaaccttctacacacccacttctctgcccagcttggagcaggacagctatcgcttt 2100
|||||
QY 2621 acaacatttgaagcctactactcttagaccaggaagtggtggtatttgcaaggtctca 2680
|||||
Db 2101 acaacatttgaagcctactactcttagaccaggaagtggtggtatttgcaaggtctca 2160
|||||
QY 2681 gcttctgtagcaggaatttgcctctcatatgagtgaggaagagcggtttaaatgctca 2740
|||||
Db 2161 gcttctgtagcaggaatttgcctctcatatgagtgaggaagagcggtttaaatgctca 2220
|||||
QY 2741 agtttctgatttggacatggggtgcggaacagtagtgcgccagacatgattttttac 2800
|||||
Db 2221 agtttctgatttggacatggggtgcggaacagtagtgcgccagacatgattttttac 2280
|||||
QY 2801 agatccagatgtaccagctctcgaggttgccttcattgattaccacagagacctctacaatc 2860
|||||
Db 2281 agatccagatgtaccagctctcgaggttgccttcattgattaccacagagacctctacaatc 2340
|||||

QY 2861 acctggagagagacagagatcgccccagcctctacgtgccccctgggttctctcaacatgt 2920
|||||
Db 2341 acctggagagagacagagatcgccccagcctctacgtgccccctgggttctctcaacatgt 2400
|||||
QY 2921 ttgctcaacagtcccgctgggattcgtagccagagtttttgatattgattttcttcagg 2980
|||||
Db 2401 ttgctcaacagtcccgctgggattcgtagccagagtttttgatattgattttcttcagg 2460
|||||
QY 2981 gaacagaggttcatttttaagtggctttaaagttctgttgggaagccataaagcccttgattc 3040
|||||
Db 2461 gaacagaggttcatttttaagtggctttaaagttctgttgggaagccataaagcccttgattc 2520
|||||
QY 3041 tgcagcatgaaacctagaacctagttgactttataaaagcgcgtacccaaccttg 3100
|||||
Db 2521 tgcagcatgaaacctagaacctagttgactttataaaagcgcgtacccaaccttg 2580
|||||
QY 3101 gcttgggtacagatgaaagaccatcaatcaggttatttgaatggacatcgtaaacagt 3160
|||||
Db 2581 gcttgggtacagatgaaagaccatcaatcaggttatttgaatggacatcgtaaacagt 2640
|||||
QY 3161 tacaagcttatgaagttgagttaccacgtccttcaagaagaacttattcgatttctctctc 3220
|||||
Db 2641 tacaagcttatgaagttgagttaccacgtccttcaagaagaacttattcgatttctctctc 2700
|||||
QY 3221 tcagtgaacacaaagaatgataaattagagaaacacacagcagcttagcgaacacaga 3280
|||||
Db 2701 tcagtgaacacaaagaatgataaattagagaaacacacagcagcttagcgaacacaga 2760
|||||
QY 3281 acctgacctcttgaacagttgcaggtggcaaatggtagatccaaagccttggaggcca 3340
|||||
Db 2761 acctgacctcttgaacagttgcaggtggcaaatggtagatccaaagccttggaggcca 2820
|||||
QY 3341 ccattgagaactcctcagcagctgagagcagctgaaagcagcagcattaccttagaac 3400
|||||
Db 2821 ccattgagaactcctcagcagctgagagcagctgaaagcagcagcattaccttagaac 2880
|||||
QY 3401 tggagcgttcggccctcgtcgcagcgttgagagagctcgcgcgagcgcagagccca 3460
|||||
Db 2881 tggagcgttcggccctcgtcgcagcgttgagagagctcgcgcgagcgcagagccca 2940
|||||
QY 3461 gcgaccgggagcctgagtgacgcagcccgagccacggcgactgacagctctgcagga 3520
|||||
Db 2941 gcgaccgggagcctgagtgacgcagcccgagccacggcgactgacagctctgcagga 3000
|||||
QY 3521 gagattgcaacacacatccacacgttccaggcctt 3555
|||||
Db 3001 gagattgcaacacacatccacacgttccaggcctt 3035
|||||

RESULT 6

AAV05886

ID AAV05886 standard; DNA; 4039 BP.

XX

AC AAV05886;

XX

DT 01-JUN-1998 (first entry)

XX

DE Human Tbc-1 gene.

XX

KW tre-2; BUB2; cdc16; mast cell; probe; hybridisation; subtraction method;
KW mouse; transcription factor; differentiation; proliferation; human; ds;
KW acute myelogenous leukaemia.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 403..3828

FT FT /*tag= a

FT ET /product= Tbc1_protein

XX

FN US5700927-A.

XX

PD 23-DEC-1997.

XX	23-DEC-1994;	94US-0363300.	
PF	XX		
XX	23-DEC-1994;	94US-0363300.	
PR	XX		
XX	(CHIL-) CHILDRENS MEDICAL CENT.		
PA	XX		
XX	Richardson P, Zon L;		
PI	XX		
XX	WPI: 1998-062437/06.		
DR	P-PSDB; AA044777.		
DR			
XX			
PT	DNA encoding Tbc1 polypeptide - useful for treating leukaemia		
PS	Claim 3; Fig 1A-B; 22pp; English.		
XX			
XX	This is the nucleotide sequence which encodes a Tbc1 (tre-2, BUB2 and		
CC	cdc16) polypeptide, isolated from a bone marrow-derived mast cell DNA		
CC	library. The screen was carried out using a probe generated by a		
CC	subtraction method which compared mRNA expression in an undifferentiated		
CC	mast cell line PB15 and cell line PG76 (PB15 cells transformed to express		
CC	the murine GATA-1 transcription factor - a factor which controls the		
CC	expression of genes involved in mast cell differentiation). Tbc1 encodes		
CC	a protein involved in the coupling of cell proliferation to cell		
CC	differentiation, which can be used to treat leukaemia (especially acute		
CC	myelogenous leukaemia) by causing leukaemic cells to differentiate.		
XX			
SQ	Sequence 4039 BP; 1042 A; 1096 C; 1060 G; 841 T; 0 other;		
XX			

Query Match	70.6%;	Score 2510.4;	DB 19;	Length 4039;
Best Local Similarity	83.7%;	Pred. No. 0;		
Matches 2949; Conservative	0;	Mismatches 536;	Indels 37;	Gaps 8;

Qy	1	atggaacaaataacattccacagcaaggaaacatctgctt-cttaacgagggtctcgtgga	59
Db	292	atggaggcaatcattccacgaggaagcatcggtttccctaaagagttctcgtgga	351
Qy	60	ttttggcctcagctggt--ggctcccctgcctgtgcatctccctgacacccatgccctg	117
Db	352	ctcgggttcagctggttggtgtctctaccogtgcattctctcaccactatgccctg	411
Qy	118	ctgcctcgggtgttggtcaggtgcgaagactcagcaggcagctccaccagaagaaacct	177
Db	412	ctgccgtgggtgtagccgaggtacaaagactcagtgccaggtgctctaaaaggagcct	471
Qy	178	gtaaccaagcaagtcoggctcttgctgtttcaacctctggactgagatgtgaacctagcca	237
Db	472	aggacaagcaagtcggcctcttggtttgcctccggactgcgtgcggctgacctg	531
Qy	238	gggagaagtcaacagtggggacccctgactctattccagatctttgagtgcaagcctcag	297
Db	532	gagaaagtcaaccaatgggaccgcctcatctgtccagatctcttgagtgcaagcctcag	591
Qy	298	cgtgttcacaaactgattcacacagtcatgacccaagttactttgcttgctgattaaag	357
Db	592	cgtgtgcacaaactgattcacacagtcacgaccgagctactttgctgctcattaaag	651
Qy	358	gaagacgctgtccacoggcagagtagtatctgctatgtgttcaaaagccatgatcaacacaaa	417
Db	652	gaggatgctgcacaggcagagcctctctgctatgtgtttaaagcagcagatcaaacaaa	711
Qy	418	gtccctgagatcatcagctccatccgtccaggcggggaagatgcccggcagagagagctg	477
Db	712	gtgcctgagatcatcagctccatccggcagggccgggaagattgcccgccaggaagagctg	771
Qy	478	cactgcgccctccgagtttcgacgcacagttttccaaagttcgaggtgctctcttcgcgcg	537
Db	772	cgttgccctccgagtttcgacgatacctcttgcacaaagtctgaggtgctctctctgtgcc	831
Qy	538	cgcgtgacggtggcgcacaaagagcttcgcgcgggcctgatgcacaggtgcacgcagag	597
Db	832	cgaagtactgtgctcacaaagagcccccacccqactgattgaacagtgatccagaaag	891

QY	598	ttcaatcacgtcagcggcagccgagggttcgagagccccgcgcaccccaacccgcccatgccc	657
DB	892	ttcaaccatgtgagctgtggtcgcaaaaggactgggaagcgcaccacggcgagccatca	951
QY	658	gcgccacaggagccagagacgtgtgcagagcccatgcgaagtctcttctccagccc	717
DB	952	gc-----gcctggcccgagggcccatgcgcaaatctctctccagcct	993
QY	718	ggcctgcgctgcctggcctttagaaaggagctgcaggatggggggcctccgaagcagcgc	777
DB	994	ggactgcgctgcctggccttcaggaaaggagtctccaggagcctagcctccgagtagcacc	1053
QY	778	ttcttcagctcctctcagagagagcagcatttgaaacacacctcattagcgcgacaataatt	837
DB	1054	tt-----tagctcctttgacaatgacatagaaagacacctcatcggggggcacaaatgctg	1107
QY	838	gtgcagcccacagatctcgagggaaaaatcgaactatgctcttcacgatgtggccagtcctgaa	897
DB	1108	gttcagcccacagacatggaggagaaccgaactatgtcttcaggtatggcccatctgaa	1167
QY	898	gtttacctatcagctcctgcagaccccaaaaaatagcaattggagaaaaattttaaggagata	957
DB	1168	gtttacctatcagctcctgcagaccccaaaaagattgcactggagaaaaattttaaggagata	1227
QY	958	tccctttgctctcaggcatcacagacgttggaccactttgggtttatctgtctcgggagctct	1017
DB	1228	tccctttgctctcaggcatcacagacatgtggaccactttgggtttatctgtctcggaagctgc	1287
QY	1018	tc-----cggaggtggcgctcttcattttgtctgttaagtgcttcagtgcaacaatgag	1071
DB	1288	tcgggtggcgagctggcgcttctcattttgtctgttaagtgcttcagtgcaacaatgaa	1347
QY	1072	gctctggttatgaaattatgatgaaccttgaaaacaggaccttcacggtgagcgagctgcag	1131
DB	1348	gctctggtttacgagatcatgatccttgaagcaggtcttcacggtgagctgcggcgag	1407
QY	1132	cagacagctaaggcgcgcagcccagctgtgtgagggtctgccccctgcgaagcctgcacaag	1191
DB	1408	cagacggttaaggcgcagcccagctctgtgagggtctgccccctgcgaagcctgcacaag	1467
QY	1192	ctctgtgagaggatagagggaatgaattcttccaaaaaaaactgaagactgcgaagcac	1251
DB	1468	ctctgcgaaggatagagggaatgaattcttcaaaccaaatagaactccagagcac	1527
QY	1252	ctgcagcattaaccaatcagcagcagcgagctattttgaagaggttcagaataatgaga	1311
DB	1528	ttgaccacactgaccataatcaggcagggccaccataatcagagaggttcagaaaaatgaga	1587
QY	1312	ccgagaattgagcgcagagagaatgaattgattattctttctgagatggtttatatgaa	1371
DB	1588	ccaagaaacgacgcagagagaatgaattaaattattctttctgagggtgcttatgaga	1647
QY	1372	gagaaacagaaagacacatcccatattggggagatgaagcagacatccagatggcagca	1431
DB	1648	gagaagcaaaaagacagccacactggggcgccaaagcagacactcagggtgcagca	1707
QY	1432	gagaataattggaagtgaattaccaccacagctgcacctcgatttagctagatgatgctgaaa	1491
DB	1708	gagaataattggagtgacctgcaccacagctgcagcgggttcagggttagcttgcgtgaag	1767
QY	1492	aacaaagcnaagagatatctttaacagagctcttttagaaagtattttgtcccggggtaataa	1551
DB	1768	aacgagcaaaagggtccttaacagagtccttagagagcattctgtcccggggtaataa	1827
QY	1552	gccagagccttcagggaacactccatcagctggtgactggaatgctcctgcttagtaca	1611
DB	1828	gccagagccttcagggaacactcccgccagctgtgtgactggaacgactccacttctagtact	1887
QY	1612	ttaagttaaccacgcaagagccatctgtgttgaaaaaggagccttgcacatctctgag	1671
DB	1888	ctagctaacacgacgagagagctgtccatggtgtgacaagagagccttccgctctctgag	1947

	27-AUG-1999; 93JP-0300253.	
PR	11-JAN-2000; 2000JP-0118776.	
PR	02-MAY-2000; 2000JP-0183767.	
PR	09-JUN-2000; 2000JP-0241899.	
XX	(HELI-) HELIX RES INST.	
PA	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;	
XX	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;	
PI	WPI; 2001-318749/34.	
XX	Primer sets for synthesizing polynucleotides, particularly the 5602	
PT	full-length cDNAs defined in the specification, and for the detection	
PT	and/or diagnosis of the abnormality of the proteins encoded by the	
PT	full-length cDNAs -	
XX	Claim 8; SEQ ID 18465; 2537pp + CD ROM; English.	
PS	The present invention describes primer sets for synthesising 5602.	
XX	full-length cDNAs defined in the specification. Where a primer set	
CC	comprises: (a) an oligo-dr primer and an oligonucleotide complementary	
CC	to the complementary strand of a polynucleotide which comprises one of	
CC	the 5602 nucleotide sequences defined in the specification, where the	
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination	
CC	of an oligonucleotide comprising a sequence complementary to the	
CC	complementary strand of a polynucleotide which comprises a 5'-end	
CC	sequence and an oligonucleotide comprising a sequence complementary to a	
CC	polynucleotide which comprises a 3'-end sequence, where the	
CC	oligonucleotide comprises at least 15 nucleotides and the combination of	
CC	the 5'-end sequence/3'-end sequence is selected from those defined in	
CC	the specification. The primer sets can be used in antisense therapy and	
CC	in gene therapy. The primers are useful for synthesising polynucleotides,	
CC	particularly full-length cDNAs. The primers are also useful for the	
CC	detection and/or diagnosis of the abnormality of the proteins encoded by	
CC	the full-length cDNAs. The primers allow obtaining of the full-length	
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and	
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to	
CC	AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632	
CC	represent oligonucleotides, all of which are used in the exemplification	
CC	of the present invention.	
XX	Sequence 2362 BP; 657 A; 581 C; 578 G; 546 T; 0 other;	
SQ		
	Query Match 59.4%; Score 2112; DB 22; Length 2362;	
	Best Local Similarity 100.0%; Pred. No. 0;	
	Matches 2112; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
Qy	1444 agtgattaccacccagtcgcactgcatgttagcttagatcgtcgataaaacaagaagcaag 1503	
Db	1 agtgattaccacccagtcgcactgcatgttagcttagatcgtcgataaaacaagaagcaag 60	
Qy	1504 agatctttaacagagtcttgaagaagtattttgccgggtaataaaggccaggacctg 1563	
Db	61 agatctttaacagagtcttgaagaagtattttgccgggtaataaaggccaggacctg 120	
Qy	1564 caggaaactccatcagtgatggaatcgtgatagctccctgcttagtacattaagtaaacacc 1623	
Db	121 caggaaactccatcagtgatggaatcgtgatagctccctgcttagtacattaagtaaacacc 180	
Qy	1624 agcaaagagccatctgtgtgtaaagaggagccttgcctatctctgagagctcctttaag 1683	
Db	181 agcaaagagccatctgtgtgtaaagaggagccttgcctatctctgagagctcctttaag 240	
Qy	1684 ctctcggctcctcgaggagacctgtcccagtgactcgagagctatccccagaagagcca 1743	
Db	241 ctctcggctcctcgaggagacctgtcccagtgactcgagagctatccccagaagagcca 300	
Qy	1744 gctccgctgtgcgccagcaggccttcagggaggagcgaacacccctgagctcctcccc 1803	
Db	301 gctccgctgtgcgccagcaggccttcagggaggagcgaacacccctgagctcctcccc 360	


```
|||||
Db 1441 cccagctctacgtgccccctgggttctctcaacatgtttgctctcacagttccccgctggga 1500
Qy 2944 ttcgtagccagagctctttgatgatattttcttcagggaacagaggttcattttaagtg 3003
Db 1501 ttcgtagccagagctctttgatgatattttcttcagggaacagaggttcattttaagtg 1560
Qy 3004 gctttaagctctgttggaagacataaagcccttgattctgcagcatgaaacccctagaacc 3063
Db 1561 gctttaagctctgttggaagacataaagcccttgattctgcagcatgaaacccctagaacc 1620
Qy 3064 atagttgactttataaaagacagctaccacaccttgcttggttacagatggaaagacc 3123
Db 1621 atagttgactttataaaagacagctaccacaccttgcttggttacagatggaaagacc 1680
Qy 3124 atcaatcaggtatttgaaatggacatcgctaaacagttacaagctttatgaagttgagtac 3183
Db 1681 atcaatcaggtatttgaaatggacatcgctaaacagttacaagctttatgaagttgagtac 1740
Qy 3184 cagctcttcaagagaacttatcgattctctctctctctcagtcagtcagacacaaagaatgat 3243
Db 1741 cagctcttcaagagaacttatcgattctctctctctctcagtcagtcagacacaaagaatgat 1800
Qy 3244 aaattagaaaaaacacacagcagcttacgcaaacagacaccttgacctctcttgaacagttg 3303
Db 1801 aaattagaaaaaacacacagcagcttacgcaaacagacaccttgacctctcttgaacagttg 1860
Qy 3304 caggtgcaaatgtagatccaaagccttgaggccaccattgagaagctctctgagcagtg 3363
Db 1861 caggtgcaaatgtagatccaaagccttgaggccaccattgagaagctctctgagcagtg 1920
Qy 3364 gagagcaagctgaagcagggcattctacattagaactgagcggctcgccctgctcag 3423
Db 1921 gagagcaagctgaagcagggcattctacattagaactgagcggctcgccctgctcag 1980
Qy 3424 acgdtgagagctgcggcgagcgagcagagcccgagcagccgggagcctgagtcgacg 3483
Db 1981 acgdtgagagctgcggcgagcgagcagagcccgagcagccgggagcctgagtcgacg 2040
Qy 3484 cagcccgagccacggcgagctgacagctctgagggagagattgcaacacccatccacac 3543
Db 2041 cagcccgagccacggcgagctgacagctctgagggagagattgcaacacccatccacac 2100
Qy 3544 gtccaggcctt 3555
Db 2101 gtccaggcctt 2112

RESULT 8
AAS34787
ID AAS34787 standard; cDNA; 1422 BP.
XX
AC AAS34787;
XX
DT 04-DEC-2001 (first entry)
XX
XX cDNA encoding novel human neoplastic disease associated polypeptide #21.
XX
XX Human; neoplastic disease associated polypeptide; cancer; gene therapy;
KW hyperproliferative disorder; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW neuroprotective; cytostatic; anti inflammatory; vasotropic; ss.
XX
OS Homo sapiens.
XX
XX WO20015163-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01358.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
```


QY 3515 gcaggagattgaacacacatccacactgtccaggcctt 3555
|||||
Db 1158 gcaggagattgaacacacatccacactgtccaggcctt 1198
|||||

RESULT 10
AAS34785
ID AAS34785 standard; cDNA; 1269 BP.
XX
AC AAS34785;
XX
DT
XX 04-DEC-2001 (first entry)
XX
DE cDNA encoding novel human neoplastic disease associated polypeptide #19.
XX
DE Human; neoplastic disease associated polypeptide; cancer; gene therapy;
KW hyperproliferative disorder; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW neuroprotective; cytostatic; anti inflammatory; vasotropic; ss.
XX
XX Homo sapiens.
XX
XX WO200155163-A1.
XX
PN
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01358.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.

PD 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 1; SEQ ID 1285; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 883 BP; 245 A; 221 C; 219 G; 193 T; 5 other;
SQ

Query Match 22.1%; Score 785; DB 22; Length 883;
Best Local Similarity 96.7%; Pred. No. 9.9e-204;
Matches 843; Conservative 0; Mismatches 23; Indels 6; Gaps 4;
QY 1444 agtgaattaccacccagtcgacatctgatttagctagatgctgaaacaaagcaag 1503
DB 1 agtgaattaccacccagtcgacatctgatttagctagatgctgaaacaaagcaag 60
QY 1504 agatctttaacagagcttttagaagaattttgtccgggggtaataaagccagagccctg 1563
DB 61 agatctttaacagagcttttagaagaattttgtccgggggtaataaagccagagccctg 120
QY 1564 caggaaacacccatcagtggtggtatcggtatgctcctgctagctacattagtaacacc 1623
DB 121 caggaaacacccatcagtggtggtatcggtatgctcctgctagctacattagtaacacc 180
QY 1624 agcaagagacatctgtgtgtgaaaggagccctgcccattctctgagagctctttaag 1683
DB 181 agcaagagacatctgtgtgtgaaaggagccctgcccattctctgagagctctttaag 240
QY 1684 ctctcggctctctcgagagacctgtccagtgactcgagagagctatctccagaagagcca 1743
DB 241 ctctcggctctctcgagagacctgtccagtgactcgagagagctatctccagaagagcca 300

QY 1744 gctccgctgtcgcgccagcaggccttccagagagcgagcgaacacccctgagtcactcccc 1803
DB 301 gctccgctgtcgcgccagcaggccttccagagagcgagcgaacacccctgagtcactcccc 360
QY 1804 atcgaatgcaggaacacctccaaacctgcccgggggtcccccgggggttttcgcaaaagaaa 1863
DB 361 atcgaatgcaggaacacctccaaacctgcccgggggtcccccgggggttttcgcaaaagaaa 420
QY 1864 ctatgaggtatcactcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1923
DB 421 ctatgaggtatcactcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 480
QY 1924 aaagcaaacacctctgtgtgattctgtgtgagtcagtcagtcagtcagtcagtcagtcagtcagtc 1983
DB 481 aaagcaaacacctctgtgtgattctgtgtgagtcagtcagtcagtcagtcagtcagtcagtcagtc 540
QY 1984 aggcagcagatattctccagtagtagccaccccgagaaagcgtagtcagtcagtcagtcagtcagtc 2043
DB 541 aggcagcagatattctccagtagtagccaccccgagaaagcgtagtcagtcagtcagtcagtcagtc 600
QY 2044 tatgaagattattcagagctggagagcttccccccacgacatctctcttagaaccagttgt 2103
DB 601 tatgaagattattcagagctggagagcttccccccacgacatctctcttagaaccagttgt 659
QY 2104 gaagatgggcccctttgccccccacgagagaaagaaagacacatctcgtgagctccga 2163
DB 660 gaagatgggcccctttgccccccacgagagaaagaaagacacatctcgtgagctnca 719
QY 2164 gaggc-tgtgtgcaaaaggctatttcttcaacagatctcgtcttagaatggagagaaagaaa 2222
DB 720 asgcttgtggcaaaaggctatttcttcaacagatctcgtcttagaatggagagaaagaaa 779
QY 2223 tcagaagctccagcc---tcgaaaatgatttgcgaaacagcgc-lgaagctcgatt 2278
DB 780 tcagaagctccagcc---tcgaaaatgatttgcgaaagtgattgctgaaacagcgc-lgaagctcgatt 839
QY 2279 atgaagaattactctcctctcttaagaagta 2310
DB 840 ntgaagaattactctcctctcttaagaagta 871
RESULT 12
AAK92006
ID AAK92006 standard; cDNA; 849 BP.
XX AAK92006;
AC AAK92006;
XX 06-NOV-2001 (first entry)
XX Human cDNA 5'-end sequence, SEQ ID NO: 466.
XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX Homo sapiens.
XX EP1130094-A2.
XX 05-SEP-2001.
XX 07-JUL-2000; 2000EP-0114089.
XX 08-JUL-1999; 99JP-0194486.
XX 11-JAN-2000; 2000JP-0118774.
XX 02-MAY-2000; 2000JP-0183765.
XX (HELI-) HELIX RES INST.
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI: 2001-524255/58.
XX


```

OY 2855 acaatcacctggagagacacgagatcgcccccagcctctacgctgcccctggttctcca 2914
DB 498 acaatcacctggagagacacgagatcgcccccagcctctacgctgcccctggttctcca 557
OY 2915 ccattgttgctccacagttcccgctgggattgtagccagagctctttgatattttc 2974
DB 558 ccattgttgctccacagttcccgctgggattgtagccagagctctttgatattttc 617
OY 2975 ttccagggaacagagctattttaagtggctttaagtctgttggaagccataagccct 3034
DB 618 ttccagggaacagagctattttaagtggctttaagtctgttggaagccataagccct 677
OY 3035 tgattctgcagcatgaaacacctagaaacccatagttgacttttataaaagcacgctaccca 3094
DB 678 tgattctgcagcatgaaacacctagaaacccatagttgacttttataaaagcacgct-ccca 736
OY 3095 accttggcttggtacagatggaaaagaccatcaatcaggta-tttgaaatggacatc-gc 3152
DB 737 accttggcttggt-acagatggaaaagaccatcaatcaggta-tttgaaatggacatcgggt 795
OY 3153 taaacagttacaagcttatg-aagttagtacca-cgtccttcaagaagaactt 3204
DB 796 taaacagttacaagcttatg-aagttagtacca-cgtccttcaagaagaactt 849

RESULT 14
AAZ50904
ID AAZ50904 standard; DNA; 17590 BP.
XX
AC AAZ50904;
XX
DT 31-MAY-2000 (first entry)
XX
DE Human TBC-1 partial genomic DNA comprising 5' end sequence.
XX
KW TBC-1; human; biallelic marker; chromosome 4; cell cycle regulator; SNP;
KW Single nucleotide polymorphism; tissue differentiation; prostate cancer;
KW linkage analysis; genetic map; detection; diagnosis; genotyping;
KW transgenic animal; screening; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_signal 1..2000
FT exon /*tag= a
FT /*note= "5' Regulatory region"
FT 2001..2077
FT /*tag= b
FT /*number= 1
FT intron 2078..12739
FT /*tag= c
FT /*number= 1
FT misc_feature 9391..9845
FT /*tag= d
FT /*note= "Amplicon 99-430"
FT primer_bind 9391..9408
FT /*tag= e
FT /*bound_moiety= "Primer B1"
FT primer_bind 9475..9493
FT /*tag= f
FT /*bound_moiety= "Primer D1"
FT misc_binding 9482..9506
FT /*tag= g
FT /*bound_moiety= "Probe P1"
FT /*note= "Detection of Biallelic marker 99-430-352"
FT 9494
FT misc_feature
FT /*tag= h
FT /*note= "Ambiguity base 'R' corresponds to 'A' in allele-1
FT and 'G' in allele-2 of biallelic marker 99-430-352"
FT complement (9495..9513)

```

```

FT /*tag= i
FT /*bound_moiety= "Primer E1"
FT /*note= "Microsequencing of marker 99-430-352"
FT complement (9828..9845)
FT primer_bind
FT /*tag= j
FT /*bound_moiety= "Primer C1"
FT /*note= "Amplification of amplicon 99-430"
FT 12292..12373
FT /*tag= k
FT /*number= 1 bis
FT intron 12374..12739
FT /*tag= l
FT /*number= 1 bis
FT exon 12740..13249
FT /*tag= m
FT intron 13250..17590
FT /*tag= n
FT /*number= 2
PN WO200008209-A2.
XX
XX 17-FEB-2000.
PD
XX
XX 06-AUG-1999; 99WO-IB01444.
XX
XX 07-AUG-1998; 98US-0095653.
XX
XX (GEST ) GENSET.
XX
XX Blumenfeld M, Bougueleret L, Chumakov I;
XX WPI; 2000-205736/18.
XX
XX New isolated human TBC-1 nucleic acids, useful for developing products
XX for the diagnosis and treatment of disorders involving cell
XX proliferation, particularly prostate cancer -
XX Claim 1; Page 93-100; 166pp; English.
XX
XX The present sequence is the partial genomic DNA of human TBC-1 gene,
XX comprising the 5' regulatory region, exons 1, 1bis and 2. TBC-1 gene is
XX mapped to a candidate region of prostate cancer on chromosome 4. Single
XX nucleotide polymorphism (SNP) is located within the biallelic marker
XX region 99-430-352, localised in intron 1 of TBC-1 genomic DNA.
XX TBC-1 gene is involved in the regulation of cell cycle and tissue
XX differentiation in mammals. An alteration of TBC-1 sequence may be
XX associated with a pathological condition, resulting in abnormal cell
XX proliferation leading to cancer, e.g. prostate cancer. The biallelic
XX markers can be used for generation of genetic maps, linkage analysis and
XX association studies. TBC-1 sequence can be used for detection,
XX diagnosis, genotyping, production of transgenic animals and screening
XX of compounds for use in therapy.
XX
XX Sequence 17590 BP; 4760 A; 3776 C; 4104 G; 4919 T; 31 other;
SQ

```

```

Query Match 11.9%; Score 422.6; DB 21; Length 17590;
Best Local Similarity 95.8%; Pred No. 5, 4e-104;
Matches 434; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

OY 1 atggaaccaataacattcacagcaaggaaacatctgcttcttaacagaggtctcgttgat 60
DB 12833 atggaaccaataacattcacagcaaggaaacatctgcttcttaacagaggtctcgttgat 12892
OY 61 ttggcctgcagctggtgggtccctgctgctgctgctgctgctgctgctgctgctgctg 120
DB 12893 ttggcctgcagctggtgggtccctgctgctgctgctgctgctgctgctgctgctgctg 12952
OY 121 ccttgggtgtggtgaggtgcgaagactcagcaggtcctcaccagaaagaaactgta 180
DB 12953 ccttgggtgtggtgaggtgcgaagactcagcaggtcctcaccagaaagaaactgta 13012

```

```
QY 181 accaagcaagtcggttgcgttccacctctgactgagatgagactgagccagg 240
Db 13013 accaagcaagtcggttgcgttccacctctgactgagatgagactgagccagg 13072
QY 241 agaagtaacagtggtgacccctgatctattccagcatctttgagtgcaagcctcagcgt 300
Db 13073 agaagtaacagtggtgacccctgatctattccagcatctttgagtgcaagcctcagcgt 13132
QY 301 gtccacaaactgattccacacagtcacgaccacgaagttacttctgtctgataaggaa 360
Db 13133 gtccacaaactgattccacacagtcacgaccacgaagttacttctgtctgataaggaa 13192
QY 361 gacgctgtccaccgagagatctctgtatgtgttcaaaagccgatgatcaacaaagtg 420
Db 13193 gacgctgtccaccgagagatctctgtatgtgttcaaaagccgatgatcaacaaagtg 13252
QY 421 cctgagatcatcagctccatcctgacggcgggg 453
Db 13253 agtgagatggagatcccaaaagactaagtggtg 13285

RESULT 15
AAZ15229
ID AAZ15229 standard; cDNA; 696 BP.
XX
AC AAZ15229;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:3698.
XX
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9338972-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01619.
XX
PR 03-APR-1998; 98US-0030666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
PA (CHIR ) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LF;
XX
WPI; 1999-494092/41.
XX
Novel human genes and their expression products which are
differentially expressed in different cell types
XX
Claim 1; Page 1311; 2479pp; English.
```

The present invention describes a library of human polynucleotides comprising the sequences given in AAZ12532 to AAZ1779. Also described is a method of detecting differentially expressed genes correlated with the cancerous state of a mammalian cell, comprising detecting at least one differentially expressed gene product in a test sample from a cell suspected of being cancerous, where the gene product is encoded by one of the 5248 polynucleotide sequences given in AAZ12532 to AAZ1779. The polynucleotides can be used as a source of primers and probes, which can

CC be used for a variety of purposes, e.g. detection of expression levels, mapping, tissue typing or profiling, forensics, genetic analysis and detection of polymorphisms. Polypeptides encoded by the polynucleotides can be used for raising antibodies for experimental, diagnostic and therapeutic purposes. The polynucleotides may also be used to construct arrays for diagnostics (which may be used to determine function of an encoded protein); and to detect differences in expression levels between two cells (e.g. to identify abnormal or diseased tissue in a human, to identify a genetic predisposition or susceptibility to a disease such as cancer). The polynucleotides of the invention are especially used in the diagnosis, prognosis and management of colorectal cancer, breast cancer, and lung cancer. The polynucleotides can also be used to screen for peptide analogues and antagonists.

XX Sequence 696 BP; 189 A; 157 C; 178 G; 142 T; 30 other;

Query Match 10.2%; Score 362; DB 20; Length 696;
Best Local Similarity 100.0%; Pred. No. 2.8e-88;
Matches 362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3194 aagaagaacttatcgtatctctctcctcagtgacacacaaagaatgataaataagaga 3253
Db 59 aagaagaacttatcgtatctctcctcctcagtgacacacaaagaatgataaataagaga 118

QY 3254 aaaccaacagcagcttacgcaaaacagacacttgacctcttgacacagttgcagtgacaa 3313
Db 119 aaaccaacagcagcttacgcaaaacagacacttgacctcttgacacagttgcagtgacaa 178

QY 3314 atggtaggatccaaagccttgaggccaccattgagaagctcctgagcagtgagagcaagc 3373
Db 179 atggtaggatccaaagccttgaggccaccattgagaagctcctgagcagtgagagcaagc 238

QY 3374 tgaagcaggccatgcttaccttagaactgagcgggtcgccctgctgcagacagtgaggg 3433
Db 239 tgaagcaggccatgcttaccttagaactgagcgggtcgccctgctgcagacagtgaggg 298

QY 3434 agctgcggcgaggagcagagagccagcagcgggagcctgagtgacacagcccgagc 3493
Db 299 agctgcggcgaggagcagagagccagcagcgggagcctgagtgacacagcccgagc 358

QY 3494 ccacggcgagctgacagcagctctgcaggagagattgcaacacacatccacactgtccaggcc 3553
Db 359 ccacggcgagctgacagcagctctgcaggagagattgcaacacacatccacactgtccaggcc 418

QY 3554 tt 3555
Db 419 tt 420

Search completed: September 9, 2002, 19:18:41
Job time: 17499 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2002, 14:24:47 ; Search time 8477.27 Seconds
(without alignments)
8775.683 Million cell updates/sec

Title: US-09-762-311-3_COPY_171_3725
Perfect score: 3555
Sequence: 1 atggaaccaataacattcac.....tccacactgtccaggcctt 3555

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: gb_vl:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_to:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match Length	DB ID	Description
------------	--------------------	-------	-------------

ALIGNMENTS

RESULT 1

BC014529

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 3023)

Strausberg,R.

Direct Submission

Submitted (24-SEP-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cdna Library Preparation: Rubin Laboratory

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

BC014529.1

BC014529

Homo sapiens

3023 bp

mrna

linear

PRI 19-DEC-2001

IMAGE:3843156, mrna.

GI:17939551

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 14 Row: C Column: 16
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES

source Location/Qualifiers
1..3023
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3843156"
/issue_type="Placenta, choriocarcinoma"
/clone_lib="NIH MGC-21"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
BASE COUNT 883 a 712 c 740 g 688 t
ORIGIN

Query Match
Best Local Similarity 77.4%; Score 2751; DB 9; Length 3023;
Matches 2751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 805 attgagaccactcattagggacacaaatttgacagccacagatcggagaaat 864
Db 10 ATTGAGAACCCCTCATTAGCGGACACAATATTGTGACGCCACAGATATCAGGAAAAT 69
QY 865 cgaactatgctcttcacgattggcagctgagtttaacctatcagctcagaccacaa 924
Db 70 CGAATATGCTCTTCAGGATGGCCAGCTGAAGTTTACCTCATCAGTCCTGACACAA 129
QY 925 aaaaatgacatggagaaataatttaagagagatactcttctcagggagtgccgagc 984
Db 130 AAAATAGCATTTGGAGAAAATTTTAAAGAGATATCTCTTGTCTCAGGCGCATCAGAC 189
QY 985 gtgaccacttggtggttattctgtcgggagctcttcggagtgccgagtgccgagtg 1044
Db 190 GTGACCACTTTGGGTTATCTCTCGGGAGTCTCGGAGGTGGCGCTTTCATTGTGTC 249
QY 1045 tgtacgtgttcagtcgacaaatgagctctggttgatgaattatgatgacctgaaa 1104
Db 250 TGTACGTGTTTCAGTGACAAATGAGGCTCTGTTGATGAATTTATGATGACCTGAAA 309
QY 1105 caggcttcacggtggccgagtcagcagacagcagcagcagcagcagcagcagcagc 1164
Db 310 CAGGCTTCAGGTGGCGCAGTGCAGCAGACAGCTAAGGCGCCAGCCAGCTGTGTGAG 369
QY 1165 gctgccccctgcaagcctgcacaaagctctgtgagagatagagggaattcttcc 1224
Db 370 GGCTGCCCTCGCAACCCCTGCACAAGCTCTGTGAGAGGTAGAGGAATGAATCTTCC 429
QY 1225 aaaaactagaaactgcaaaagcctgacgacattacaaatcagagcagcagcagc 1284
Db 430 AAAACAAAATAGAACTGCAAAAGCACTGACGACATTAACCAATCAGGAGCAGCGACT 489
QY 1285 atttttgaagaggttcagaattgagaccagaaatgagcagcagcagcagcagcagc 1344
Db 490 ATTTTGAAGAGGTTTCAGAAATTTGAGCCGAGAAATAGCAGCGCAGAGAGAAATGATT 549
QY 1345 attcttttctgagatgtttatatgagagaaacagaaagacacattcattggggag 1404
Db 550 ATTTCTTTTCTGAGATGTTTATATGAAGAAACAGAAAGACATCCATATTGGGGAG 609
QY 1405 atgaagcagacatcgagatggcagcagcagcagcagcagcagcagcagcagcagc 1464
Db 610 ATGAGCAGACATCGCAGATGGCAGCAGAAATATTGGAATGAATTAACCAACCCAGTGC 669

QY 1465 actcgatttaggctagatatgctgaaaaacaaagcagaagagatctttaaagagctcttta 1524
Db 670 ACTCGATTAGGCTAGATATGCTGAAAAACAAGCAAGAGATCTTTAACAGAGTCTTTA 729
QY 1525 gaaagtatttttcccggggtaataaagccagagggcctcaggaacacacccatcagtg 1584
Db 730 GAAAGATATTTGTCGCCGGGTAAATAAGCCAGAGCCCTGCAGAACACATCCATCAGTGTG 789
QY 1585 gatctggatagctccctgtctagtagtatttaagtaacacacagcagaagagcctctgtgt 1644
Db 790 GATCTGATAGTCCCTGTGTAGTACATTAACTAACACACAGCAAGAGCCATCTGTGTG 849
QY 1645 gaaaagaggccttgcccatctctgagagctctttaaagctctcctcggtcctcctggagac 1704
Db 850 GAAAAGGAGGCCCTTGCCCATCTCTGAGAGCTCTTTAAGCTCTCTCGGCTCTCTCGGAGAC 909
QY 1705 ctgtccagtgactcggagagtgatctcccaagagccagctccgctgtcgcgccacag 1764
Db 910 CTGTCCAGTGACTCGAGAGTCTATCTCCAGAAAGAGCCAGCTCCCTGTCTGCCCCAGCAG 969
QY 1765 gccttcaggaggcgaacaaacccctgagtcacttcccccatcgaatgccaggaacctcca 1824
Db 970 GCTTTCAGGAGCGGAGCAAAACACCTGTGAGTCACTTCCCATCGAATGCCAGAACCTCCA 1029
QY 1825 caactctcccggggtcccggggtttcgaagggaaacttatgaggtatcactcagtg 1884
Db 1030 CAACCTGCCCCGGGGTCCCGGGGTTCGCAAAAGGAACTTATGAGGTATCACTCAGTG 1089
QY 1885 agcacagagcgcctcatgaacgaagagctttgaatccaaagcaacacatctctgtgat 1944
Db 1090 AGCACAGAGCCCTCATGAACGAAGGACTTTGAATCCAAAGCAAAACCATCTTGTGTAT 1149
QY 1945 tctgtgtgactcctgtgaagaccggagcttctcctcagagatgaagattatcagagctg 2004
Db 1150 TCTGTGGGACTCTGTGAAGACCCGGAGGCACTTCTCGAGGCAGCAGATATTCCTCGCA 1209
QY 2005 gtagcaccgccgagaaggcgtcgattcttcagcagatgaagattatcagagctg 2064
Db 1210 GTAGCCACCCCGAAGAGCGTGGCATCTTCAGCAGATATGAAGATTATTCAGAGCTG 1269
QY 2065 gtaggactctcccccagatctctttagaacctgttggaagatgggccccttggcccc 2124
Db 1270 GGAGAGCTTCCCCAGATCTCCTTTAGAACCAGTTTGTGAAGATGGGCCCTTTGGCCCC 1329
QY 2125 ccacagaggaaaagaaaggacatctcgtgagctccgagagctgtggcaaaaggctatt 2184
Db 1330 CCACCAGAGGAAAAGAAAGGACATCTCTGAGCTCCGAGAGCTGTGGCAAAAGGCTATT 1389
QY 2185 cttaacagatactgctgttagaatggagaggaataatcagaagctccaagcctctgaa 2244
Db 1390 CTTCAACAGATACTGTCTTAGAATCGAAGGAAATCAGAAGCTCCAGAGCTCTGAA 1449
QY 2245 aatgattgtctgaacagcctgaagctgattgaagaaattactcctcctctcttaa 2304
Db 1450 AATGATTGCTGAACAGGCCCTGAAGCTCGATTATGAAGAAATTACTCCTCTCTTAA 1509
QY 2305 gaagtaactacagtggtggaaaagatgcttagcactccaggagatcaaaaaattagttt 2364
Db 1510 GAAGTAACATACAGTGTGGAAAAGATGCTTAGCATCTCAGGAGAGATCAAAAATTAGTTT 1569
QY 2365 gacatggaaaataatgactcgtggtgtggcaaggtgtgcacgtcatcaccgaggtgaa 2424
Db 1570 GACATGAAAATAATGCACCTCGCTGTTGGGCAAGGTGTGCCACGTCATCACCAGGTTAA 1629
QY 2425 atctggaaattctcagctgagcaattccaccttaaacaccagtttccaccagaaacagcag 2484
Db 1630 ATCTGGAATTTCTAGCTGAGCAATTCACCTTAACACACAGTTTCCCGACCAACAGCAG 1689
QY 2485 ccaaaaggaatgctccatcaaaagaactcttaagcagctgacttcccagcagatcgatt 2544
Db 1690 CCAAGGATGTCCCATACAAAGACTCTTAAAGCAGCTGACTTCCAGCAGCATGCGATT 1749

QY 778 ttcttcagctccttcgagggagcgcacattgagaaacacacctcattagcggacacaaatt 837
Db 1054 TT-----TAGCTCTTGTGACATGACATAGAGAACACCACTCATCGGTGGGCAATGTG 1107
QY 838 gtccagccacagatatcgaggaaatcgaaactatgtctcttcacgattggccagtgtaa 897
Db 1108 GTTCAGCCCAACAGACATGGAGGAGAACCGAACTATGCTGTTCAGGATTGGCCCATCTGAA 1167
QY 898 gtttacctcatcagtcctacacacaaataatagcattggagaaaaattttaaggagata 957
Db 1168 GTTTACTCTCATAGTCTCTGACACCAAAAGATTGCACTGGAGAGAAATTTAAGGAGATA 1227
QY 958 tccttttgcctcagggccatcacacagctggaccactttgggttttatctgtcgggagtct 1017
Db 1228 TCCTTTTGTCTCTCAGGGCATACACATGTGGACCATTGTGGATTTCCTGCCGAGAGTGC 1287
QY 1018 tc-----cggaggtggcggtcttcattttgtctgttaagtgtttcagtcgacaaatgag 1071
Db 1288 TCGGGTGGCGGAGTGGCGGCTTTTCATTTTGTCTGTACGTGTTCAGTGCACAAATGAA 1347
QY 1072 gctctggttgatgaataattatgatgacctgaacacagccttcacggtggccgagtgca 1131
Db 1348 GCTCTGTTGACAGATCATGATGACTCTGAACAGCGCTTTCAGGTAGCTGGGTGCGAG 1407
QY 1132 cagacagctaaggcgcacccagctgtgtgaggggtgccccctgcaaaagcctgcacaa 1191
Db 1408 CAGACGGCTAAGGCACAGCCAGCTCTGTGAGGGCTGCCCTTTCGAAGGCTGCAACAAG 1467
QY 1192 cctctgagaggatagaggaatgaattcttccaaacaaactagaaactgcaaaagcac 1251
Db 1468 CTCTCGAAGGATAGAGGAATGAATTCATCTAAACCAAAATTAGAATCCCAAGAGCAC 1527
QY 1252 ctgacacataaccacacagcagcagcgcactatttttgaagaggtttcagaaattgaga 1311
Db 1528 TTGACCACACTGACCAATCAGGAGCAGCCACCATTATTCGAGGAGGTTTCAGAAATTGAGA 1587
QY 1312 ccgagaaatgagcagcaggaatgaattgatttatttcttctctgagattgtttatatgaa 1371
Db 1588 CCAAGAAACAGCAGCAGAGAGAAATGAATTAATTAATTTCTTCTGAGGTGCTATATGAA 1647
QY 1372 gagaaacagaaagacacacatcatttgggagagatgaacacacatcgagatggcagca 1431
Db 1648 GAGAACCAAAAGAGCAGCAGCCACCTAGCTGAGCGGTTCAGGTTAGATTTCGCTGAAG 1767
QY 1432 gagaatttgggaagtgaattacacccagctgcacactcgttaggttagatgatatgtgaaa 1491
Db 1708 GAGAATATTGGGAGTGACCTGCGCACCCAGCTGCTAGCGGTTTCAGGTTAGATTTCGCTGAAG 1767
QY 1492 acaaaagcaagagatcctttaacagagcttttagaaagtattttgtcccggggtaataaa 1551
Db 1768 AACAGAGCAAAAGAGGTTCCTTAACAGAGTCCCTTAGAGAGCATTTCTGTCCCGGGTAATAA 1827
QY 1552 gccagagcctcaggaacacatccatcagctggtgattctgtagatgctcctctctagtaca 1611
Db 1828 GCCAGAGGCTCAGAGACCATTCGCGAGTGGATCTGGACAGTCCACTTCTTAGTACT 1887
QY 1612 ttaagttaacaccagcaagagcactctgtgtgaaagagggccttgcceatctctgag 1671
Db 1888 CTAAGTAACACCAGCAGAGCTGTCCATGGGTGACAGGAGGCGCTTCCCGCTCTCTGAG 1947
QY 1672 agctcctttaagctcctcctcagggaccctgtccagtgactcagagagatcatctc 1731
Db 1948 ACCTCTTCAAGCTCTTGGCTTCTCAGATGACCTGTGCAGTGACTCAGAGGGCCACATT 2007
QY 1732 ccagaagagcagctccgctgtgcgcccagcagggccttcagggagcagagcaaaacccctg 1791
Db 2008 CGAGAAGAGTCTGCGCTGTGTCACCCAGCAGGCGGTTTCAGAAGAGAGAGCCAAACCCCTG 2067
QY 1792 agtcacttccccatcgaatgcccaggaactccacaaactgcccgggggtccccgggggtt 1851
Db 2068 AGTCATTTCCAGTAGTAGTGCCCTCGCGCTCCGAAGACCTTGCCACAGAGCTCTCCAGGGGTC 2127
QY 1852 tcgcaaaaggaaacttatgaggtatcactcagtgagcacagagacgcctcctatgatgaacgaa 1911

Db 2128 TCTCAAAAGGAACTCATGGGTACCCTCGTGAGCACAGAGAGCGCTCATGAACGCAAG 2187
QY 1912 gactttgaatccaaaacaaacacatcttctggtgattcttgggtgggactcctctggaagaccgg 1971
Db 2188 GACTTTGAATCCAAAGCAAAACCACTTGGGTGACACAGATGGGACCCCGGTGAAGACCCGG 2247
QY 1972 aggcattcctcggagcagcagatattcctccagtagtagcaccaccccgagaaaggcgtgcgat 2031
Db 2248 CGGACTCGTGGGAGACAGCAGATATTCTTCGAGTGGCCACTCCACAGAAAGGCTTTGTGAC 2307
QY 2032 tctccagcagatatgaagattattcagagctgggagagcttcccccaagatctctctta 2091
Db 2308 TCCCCGAGCAGATATCAGATATTTCGAGCTGGGAGAGCTCCTCCACGCTCCCTTTA 2367
QY 2092 gaaccaatttgtgaagatggggcccttggcccccccccagagagaaagaaaaagaaacatct 2151
Db 2368 GAACCGGTGTGAGACCGGCCATTTTGG-CCAGTACAGAGAAAGAAAGAGAGAGCTCA 2426
QY 2152 cgtgagctcgcagagctgtggcaaaaggctattcttcaacagatactg-ctgcttagaat 2210
Db 2427 CGCGAGCTTCGAGAGCTGTGGAAGAGGCATCTTCGACGAGATCTCTGCTCCTCAGGAT 2486
QY 2211 ggaagaggaaatcagaagctccaagcctctgaaatgatttctgtaacaaagcgccgtgaa 2270
Db 2487 GGAGAAAGGAAATCAGAAGCTACAAGCCTCTGAAAACGATTTGCTGAACAAACGCTCAA 2546
QY 2271 gctcgattatgaagaaattactcctcctgtttaaagaagttaactacagtggtgggaaagat 2330
Db 2547 GCTTGACTATGAAGAAATCCTCCGTGCTTAAAGAAAGTCACTACAGTGTGGGAAAGAT 2606
QY 2331 gcttagcactccaggaagatcaaaaataagtcttgacatggaaaaaatacactcggctgt 2390
Db 2607 GCTTAGCACTCCAGGAGATCCAAAATTAAGTTTGACATGGAAAGAAAGTGCATCACTGT 2666
QY 2391 tgggcaaggtgtgccacgtcatccagcaggtgaaactctggaatcttctagctgagcaatt 2450
Db 2667 TGGGCAAGGTGTGCCAGCTCATCACCGAGGTGAGATCTGGAATTTCTAGCTGAGCAGTT 2726
QY 2451 ccaccttaaaacccagtttccacagcaaaacagcagcaaaaggatgtgccaataaaagact 2510
Db 2727 CCACCTTAAACACCCATTTCTTAGTAAACAGCAGCCAAAGAGCGTGCCTTACAAAGAGCT 2786
QY 2511 cttaagcagctgacttcccagcagcagctgattcttattgaccttggcggaacctttcc 2570
Db 2787 CCTGAAGAGCTGACCTGCCAGCAGCAGCCCATTTCTATCGACCTCGGGCGGAACCTTCC 2846
QY 2571 tacacacccatacttctctgccagcttggagcagcagcagctatcgtcttaacacatttt 2630
Db 2847 AACACATCCATCTCTCTGCCAGCTTGGAGCAGGTCACTGCTCACTTTTACAACATTTCT 2906
QY 2631 gaagcctactcactcttagaccaggaagtggaatatgtccaaaggtctcagcttctgtagc 2690
Db 2907 GAAGCGCTACTCGCTCTGACCCAGGAGGTGGATTAAGTACTGCCAAGGTTCAGGTTTGTGGC 2966
QY 2691 aggcatttctctctcatatagtgaggaagagcgttttaaaatgctcaagtcttctgag 2750
Db 2967 AGGCATTTTGTCTCTCATAGTGAGGAGAGCGGTTCAGATGCTCAAGTTCCTGAT 3026
QY 2751 gttgacatggggctgcgaaacagtatcgccagacatgattattttacagatccagat 2810
Db 3027 GTTTCACATGGGGCTGCGGAAACAGTATCGGCCAGACATGATTAATTTTGCAGATCCAGAT 3086
QY 2811 gtaccagctctcagaggttgcctcatgattacacagagacacctctcaatcaactggagga 2870
Db 3087 GTACCAGCTGCACGGCTCCTCCACGATTAACACCGAGACCTCTCAACACCACTGGAAGA 3146
QY 2871 gcacagatcgcccgccagcctctacgtgccccctggttctccacactggttctgctcaca 2930
Db 3147 GCACGAGACTGGCCCCCTACGTACCGGCTCCCTGTTTCTCACCGGTGTTGCGCCTCACA 3206
QY 2931 gttcccgctgggattcgtagccagagctcttgatgatattttctcaggggaacagaggt 2990

```
Db 3207 GTTCCCACTCGGCTTTGTAGCCAGAGCTCTTTGATATGATCTTCTTCAGGGATCAGAGGT 3266
Qy 2991 catatttaaagtggcttaagtctgttgggaagccataaagcccttgatctgcagcatga 3050
Db 3267 CATATTTAAAGTAGCTTTAAAGCTTTTGGGAGCCATAAGCCCTTGATTCACAGCATGA 3326
Qy 3051 aaactagaacaacatagtagctttataaaaagcagctaccacaacctgcttggtaca 3110
Db 3327 GAACCTGGAAACCATCTGGGACCTTCATAAGAACACACTCCCAACCTGGGCTGTGTGCA 3386
Qy 3111 gatgaaaagaccatcaatcagtagtttgaaatggacatcgtaaacagtgtaacagctta 3170
Db 3387 GATGAGAAGACCATCAGTCAGTGTGTTGAGATGACATCGCAAGCAGCTCCAGCCCTA 3446
Qy 3171 tgaagttagtaccagctctctcaagaagaactatcagttactctctctctcaagtgacaa 3230
Db 3447 TGAGTTCGAGTAGTACCACTGCTCGAGGAGAGCTTATTGAGTCTCGCCTCTCAGTGACAA 3506
Qy 3231 ccaagaatggaataatagagaacacacacagcagcttacgcaacagaaaccttgacct 3290
Db 3507 CCAGAATGGAGAATTTGGAGAAAACCCACAGCAGCTTGCGCAACAGAACCTTGACCT 3566
Qy 3291 cettgaacagttgtaggtggcaaatggttaggtccaaagccttgaggccaccattgagaa 3350
Db 3567 CCTGAGCAGTTCAGGTGGCAATGCTAGGATCCAAAGCCTTGAAGCCCGCTAGAGAA 3626
Qy 3351 gctctgagcagtagagcaagcagcagcagcagcagcagcagcagcagcagcagcagc 3410
Db 3627 ACTTCTACCCAGCAGAGTAAGCTGAAGCAGCTGCGCTGACCTGGAGGTGGAGCG--T 3684
Qy 3411 ggcctgtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3470
Db 3685 CGCCTGTGTCAGATGCTGGAGAGAGCTGCGGAGGCAAGAGCGCCGCGCCAGACTCCAGA 3744
Qy 3471 gctgtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3512
Db 3745 GCAGACTGCACCCAGCTGAGGCCACAGCGGATTTGACCGCT 3786

RESULT 3
MMU33005 house mouse.
LOCUS Mus musculus tbc1 mRNA, complete cds. 4039 bp linear ROD 12-JAN-1996
DEFINITION U33005
ACCESSION U33005
VERSION U33005.1 GI:988220
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4039)
AUTHORS Richardson,P.M. and Zon,L.I.
TITLE Molecular cloning of a cDNA with a novel domain present in the
tre-2 oncogene and the yeast cell cycle regulators BUB2 and cdc16
Oncogene 11 (6), 1139-1148 (1995)
JOURNAL Oncogene 11 (6), 1139-1148 (1995)
MEDLINE 96032578
PUBMED 7566974
REFERENCE 2 (bases 1 to 4039)
AUTHORS Richardson,P.M. and Zon,L.I.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-1995) Leonard I. Zon, Hematology, Childrens
Hospital of Boston, HHMT, 300 Longwood Ave., Boston, MA 02115, USA
FEATURES
source
location/Qualifiers
1..4039
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_type="mast cells"
1..4039
/gene="tbc1"
403..3828
/gene="tbc1"
/feature="ORF1"
/codon_start=1
```

```
/product="Tbc1"
/db_xref="GI:988221"
/translation="MPMLPVWVAEVRRLSGQCSKKEPRTKQVRLWVSPSGURCBPDLE
KSQWPLDICSIFEKQRPVHKLHNSHDPYSFACLIKEDAAHQRSYCYVFKADDQT
KVEPIISSIROAGKIAEQEELRCEPDDTAKKFEVLCFGRVTVAHKAPALIDEC
LEKHNHVSQGRRTDWEAPTQPSAPGPPPMKSFOPGLRSLAFKPEFQDASLRSTF
SSFUNDIENHLIGGHNVPQPTDMEENRTMLTIGPSEVYLLSPDTKKTALXNPKFELS
FCSGIRHVDHFGFIRECSGGGPFHYCYVFCQETNEALVDEIMTLKQATVAAV
QOTAKAPALCEGCPLOGLKLCERIEGMNSKTKLEKHLTLTTNOQATIFEVQ
LDRPNQKRENELISFLRCLYEKQESHGTAPKQLQVAENIGSDLPSPASRFR
LDSLNKRAKRLSTLESILSRGNKARGLOHDSVLDLSTSTLSNTSKLSMGDK
EAPVSTSFKLIGSDSDLSSEGHAEESALLSPOAFRRRANTLHFVPEVCAPP
EPASSQVSOQKLMRYVSTFPHERKOFESKANHLGDTGTPVTRRHSRQOIF
LRVAPQKACDSPSEYEDYSELGELPRLSPLEPCVEDGPGFYQKRRRRRIASFSCG
KRSCRSCLVMEKENOKLOASENLLKRLDYEITFCLKEVITWERMULTSPG
RSKTKFDMERKVSAGVGQPRHGEINWFLAEQFHLKPPFPKQKQDVKYKELKK
LTSQHAILLIDGTFTHPYFSAQLAGQSLXNLYLILKAYSLLDQEVYCGLSFVAG
ILLHMSBEEAFKMLKFLDMGRKQVPMIILQIQYQLSRLLHHDYHRLYHLE
EHENGPTTAAFWFLTFASQFPLGFVARVDMIFLOQSEVIFKVALSLLSGHKPLIL
QHENLETIVDFIKNTLNLGLVQMEKTIQOVEMDIKQLOAYEVEYHVQEELESS
PLSDNRMKLEKNTSLRKNLDDLEQLQVANARIQSLEATVEKLLTSEKLRQAL
TLEVRPPAADGGGAEEAKRPAQHRHLPAGARRHLTAARRDCAPLTLKSP"
BASE COUNT 1042 a 1096 c 1060 g 841 t
ORIGIN
Query Match 70.6%; Score 2510.4; DB 10; Length 4039;
Best Local Similarity 83.7%; Pred. No. 0;
Matches 2949; Conservative 0; Mismatches 536; Indels 37; Gaps 8;
Qy 1 atggaacaaataacattcacagcaagaaacacatgcctt-ctaacagaggtctcggtga 59
Db 292 ATGGAGGCANTACATTCACAGCAGGAAGACATCGTTCCCTACGAAGTCTCCGTGGA 351
Qy 60 ttttgccctgcagctggt--gggctccctgctgtgcatctccctgaccaccatgccatg 117
Db 352 CTTCGGCTTGACGCTGGTGGTCTCCTTACCGCTGCACTCTCACCACTATGCCCATG 411
Qy 118 ctgacctgggtgtggtgagtgagcagcagcagcagcagcagcagcagcagcagcagcagc 177
Db 412 CTGCGCTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 471
Qy 178 gtaacaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 237
Db 472 AGGACAAACAAAGTCCGCTTTGGGTTTGGGCTCGGACTCGGCTCGGAGCTGACCTG 531
Qy 238 gggagaagtcaacagtgaggatccctgactattccagcagcagcagcagcagcagcagcag 297
Db 532 GAGAAAGTCAACCATGGAGCCGCTCATCTCTTCCAGCATCTTTCAGTGAAGCTCAG 591
Qy 298 cgtgtcaaaactgatttcaacaaagcagcagcagcagcagcagcagcagcagcagcagcag 357
Db 592 CGTGTGCACAAACTGATTACACAAAGTACACGACCCGAGCTACTTTCCTTCCTCATTAAG 651
Qy 358 gaagacgctgtccacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 417
Db 652 GAGGATGCTGCACAGCGCAGAGCCTCTGCTATGTTTAAAGACAGACAGCATCAACAAAA 711
Qy 418 gtgctgagatcatcagctccatccgctcagcagcagcagcagcagcagcagcagcagcagc 477
Db 712 GTGCTGAGATCATCAGCTCCATCCGCGCAGCGCGGAGAGATTGCCCGCCAGGAAGCTG 771
Qy 478 caactcccgctccgagtgacacacgcttttccagaagattcgagagtgctcttcgcgcg 537
Db 772 CGTTGCCCTCCGAGTTCGACGATACCTTCGCCAAAAAGTTTCAGGTGCTCTTCTGTGTC 831
Qy 538 cgcgtgacggtggcgacacaaagcagcagcagcagcagcagcagcagcagcagcagcag 597
Db 832 CGGGTGACTGTGGCTCACAAGAAAGCCGCCACCCGACCTGATTGACAGAGTGTATCGAAG 891
Qy 598 ttcaatcagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 657
```

Db 892 TTCACCATGTGAGCTGTGTCGCAGAACGGACTGGAGCGCCACCGGGCAGCCATCA 951
Qy 658 gcgcccacagggagccagcgtgtgcgagcccatgcgcaagtctcttcccagccc 717
Db 952 GC-----GCTGGCCCCAGGCCCATGGCAAAATCTTCTCACAGCCT 993
Qy 718 ggcctgcgtcgtgcctcttagaaggagcgtcagatggggcctcgaagcagcgc 777
Db 994 GGAAGTGGCTGCGTGGCTTCAGAGAGGAGTCCAGAGCCTAGCCTTCGCGATGACCC 1053
Qy 778 ttctcagctccttcgagagagcgacattgagaaacacacctctattagcgacacaaatt 837
Db 1054 TT-----TAGCTCTTGACATGACATAGAACACCTCATCGTGGGCACATGTG 1107
Qy 838 gtccagccacagatatcgaggaaaaatcgaactatgctcttcacgattggccagctgaa 897
Db 1108 GTTCAGCCACAGACATGGAGGAGAACCGAATATGCTGTTTCAAGATGGCCCATCTGAA 1167
Qy 898 gtttacctcagctcctgaacacaaaataagcattgagaaaaattttaaggagata 957
Db 1168 GTTTTACCTATCATGCTCTGACACCAAAAGATTGCACTGGAGAAAATTTTAAGGAGATA 1227
Qy 958 tctctttgctcctcagggcatcagacagctggaccactttgggtttatctgtcgtggagtct 1017
Db 1228 TCCTTTTGTCTCAGGGCATCAGACATGTGGACCCTTTGGATTCTCGCGAGAGTGC 1287
Qy 1018 tc-----cggaagtgggcgtcttcatttctgtctgttaagtggttcagtcgacaaatgag 1071
Db 1288 TCGGGTGGGGCAGTGGCGCTTTTCACTTTTGTCTGTACGTGTTCAGTGCACAAATGAA 1347
Qy 1072 gctcgtgtgatgaattatgagccctgaacagagccttcacgctggccgagtgag 1131
Db 1348 GCTCTGTTGACAGATCATGATGACTCTGAGCAGGCTTTACGCTAGTGGCGTGCAG 1407
Qy 1132 cagacagtaagcgcagccagccagctgtgtgaggggtgccccctgcaaaagcctgcacaag 1191
Db 1408 CAGACGCTAAGCACACCGCCAGCTCTGTGAGGGGTGCCCCCTTGCNAAGGCTTGCACAAG 1467
Qy 1192 ctctgtgagaggatagaggaatgaattctccaaacaaaactagaactgcaaaagcac 1251
Db 1468 CTCTCGAAAGGATAGAGGAATGAATTCATCTAAACCAAAATTAAGTCCAGAGCAC 1527
Qy 1252 ctgacgacattaaccaatcagggcagggagctatttttgaagaggttcagaaattgaga 1311
Db 1528 TTGACCACACTGACCAATCAGAGCAGCGCCACCATATTCGAGGAGGTTTCAGAAATTGAGA 1587
Qy 1312 ccgagaaatgagcagcagagaaatgaattgatttattcttctgtgagatgtttatgaa 1371
Db 1588 CCAAGAAACAGCAGCGAGAGAAATGAATTAATTAATTTCTTTTCTGAGGTGCTTATATGAA 1647
Qy 1372 gagaacagaaagacacacatccattatggggagatgaagcagacatcgagatggcagca 1431
Db 1648 GAGAGCAAAAGACAGCACAGCCACACTGGGGCCGCAAGCAGACACTACAGGTGGCAGCA 1707
Qy 1432 gagaatttggaagtgaattaccaccacagtgccactcgtatttaggtagatatgctgaaa 1491
Db 1708 GAGAATATTGGGAGTGACCTGCCACCCAGTGTAGCCGGTTACGTTAGATTGCTGTAAG 1767
Qy 1492 acaagcagaagagatctttaacagagcttttagaagatattttgtcccggggtaataaa 1551
Db 1768 AACAGAGCAAAAGAGTCTTTAACAGAGTCCCTAGAGAGCATTTGTCCCCGGGGTAATAA 1827
Qy 1552 gccagagggcctgcagaaacactccatcagtggtgattctgagatgctccctgtctagtaca 1611
Db 1828 GCCAGAGGCTCGAGACCAATTCGCCAGTGTGTGATCTGGACAGCTCCACTTCTAGTACT 1887
Qy 1612 ttaagtaacaccagcaagagccatctgtgtgtgaaaaggagggccttgcccactctgag 1671
Db 1888 CTAAGTAACACCAGCAAGAGCTGTCCATGGGTGACAAAGGAGGCTTCCCGCTCTGTGAG 1947
Qy 1672 agctcctttaagctcctcggctcctcggaggaacctgtccaagtgaactcggagagctatctc 1731
Db 1948 ACCTCCTTAAAGCTCTTGGCTCCTCAGATGACCTGTCCAGTGAATTTATTTTTCAGATCCAGAT 2007

Qy 1732 ccagaagagccagctccgctgtctgccccagcaggcccttcaggaggcgagcaaacacccctg 1791
Db 2008 GCAGAGAGTCTGCCCTGTGTTCACCCAGCAGCGTTCAGAAAGAGAGCAACACCCCTG 2067
Qy 1792 agtcacttccccatgaattgcaggaacctccacaacctgcccggggtccccgggggtt 1851
Db 2068 AGTCATTTCCAGTAGAGTCCCTGCGCTCCAGAACCTTCCCCAGAGCTCTCCAGGGTTC 2127
Qy 1852 tcgcaaggaacacttatagggtatcactcagtgagcacagagacgcctcatgaacacaaag 1911
Db 2128 TCTCAAAAGAACTCATCGGTACCACTCCGTGAGCACAGAGACGCTCATGAACCAAG 2187
Qy 1912 gactttgaatccaaaacaaacccatcttgggtattcttgggtggactcctgtgaaagaccgg 1971
Db 2188 GACTTTGAATCCAAAGCAAAACCACTGGGTGACACAGATGGGACCCCGCTGAAGACCCCG 2247
Qy 1972 aggcactcctggaggcagcagatatctcccgagtagccaccccgacagaaggcgtgcgat 2031
Db 2248 CGGCACCTCTGGAGACAGCAGATATTCCTTCGAGTGGCCACTCCACAGAAGGCTTGTGAC 2307
Qy 2032 tcttcagcagatatgaagattatttcagagctggggagagcttcccccaagatctctttaa 2091
Db 2308 TCCCCGAGCAGATATGAAGATTATTCGAGCTGGGAGAGCTCCCTCAGCTCCCTTTTA 2367
Qy 2092 gaaccagttgtgaagatggcccttggccccccaccagaggaagaaagagacatct 2151
Db 2368 GAACCGGTGTGAGACCGGCCCATTTTGG-CCAGTACAGAGAAAGAGAGAAAGAGCTCA 2426
Qy 2152 cgtgagctccgagagctgtggcaaaaggctattcttcaacagatactg-ctgctttagaat 2210
Db 2427 CGCGAGCTTCGAGAGCTGTGGAAAAAGGCCATCTTCCAGCAGATCTCGCTCGTCAGGAT 2486
Qy 2211 gagaaggaataatcagaagctccaagcctctgaaaaatgatttgcgtgaaagcgccgtgaa 2270
Db 2487 GGAGAAGGAGAAATCAGAAGCTTCAAGCCTTGAACCGATTCTGTAACAAACGCCCTCAA 2546
Qy 2271 gctcgattatgaagaaattactcctcctgtctaaagagtaactacagtgctgggaaagat 2330
Db 2547 GCTTGACTATGAAGAANTCCTCCGTGTCTTAAAGAAAGTCACTACAGTGTGGGAAAGAT 2606
Qy 2331 gcttagcactccagggaagatcaaaaaataagtttgacatggaaaaaatagcactcggctgt 2390
Db 2607 GCTTAGCACTCCAGGAAGATCCAAAAATTAAGTTTGATGGAATAAGTGCACTCAGCTGT 2666
Qy 2391 tgggcaagtggtgcacgtcatcacagagtgaaatctgaaatttctagctgagcaatt 2450
Db 2667 TGGGCAAGGTGTGCCACGTATCACCAGGTGAGATCTGGAAATTTCTAGCTGAGCAGTT 2726
Qy 2451 ccaccttaaacacccagtttcccagcaaacagcagccaaaggatggtccatacaaaagact 2510
Db 2727 CCACCTTAAACACCCCATTTCTTAGTAAACAGCAGCCAAAGGACGTGCCCTTACAAAGAGCT 2786
Qy 2511 cttaagcagctgactccocagcagatcgatctcttattgaccttggcggaacctttcc 2570
Db 2787 CCTGAAGAGAGCTGACCTCGCAGCAGCAGCCCATTCATCGACCTCGGGGAAACCTTTCC 2846
Qy 2571 tacacacccatctctctgcccagcttgagcagcagcagcagctcttatacaacatttt 2630
Db 2847 AACACATTCATCTCTCTGCCCGCTTGGAGCAGCTGAGCTGCTACTTTTACAACTTCT 2906
Qy 2631 gaaggcctactcactcttagaccaggaagtggaattatgccaaaggtctcagcttttagc 2690
Db 2907 GAAGGCTTACTCGCTTCTGGACCAGGAGGTGGATACTGCCAAGGTCTCAGCTTTGTGGC 2966
Qy 2691 aggcatttgccttctcatatgagtgaggaagggcgtttataaagtctcaagtttctgat 2750
Db 2967 AGGCATTTTGTCTTCTTACATGAGTAGGAAGAGGCGTTCAAGATGCTCAAGTTCCTGAT 3026
Qy 2751 gtttgcacatgggctgcgaaacagtatcgccagacatgattattttacagatccagat 2810
Db 3027 GTTTGACATGGGCTCGGAAACAGATATCGCCAGACATGATTTATTTTTCAGATCCAGAT 3086

Db 421 TCTGTGTGTGAAAGAGGAGGCTTTGCCATCTCTGAGAGCTCTCTTAAGCTCTCGGCTCC 480
Qy 1696 tcggaggacctgtccagtgactcgagagtcattctcccaagagagccagctccgtgtcg 1755
Db 481 TCGGAGGAGCTTCCAGTGACTCGGAGAGTCATCTCCAGAAAGAGCCAGCTCCGCTGTCG 540
Qy 1756 cccagcagccttcagagagcgagcaaacacctgagtcacttcccatcaatgagcag 1815
Db 541 CCCCAGAGCCCTTCAGGAGGCGAGCAACACCTTGAGTCATTTCCCNATCGAATGCCAG 600
Qy 1816 gaacctccaaacctccgggggtcccccgggggttgcgaagaaacttatgaggtat 1875
Db 601 GAACCTCCCAACCTCCCGGGGGTCCCGGGGGTTCCGAAGAAACTTATGAGGTAT 660
Qy 1876 cactcaagtacagacagagcgcctcatgaacgaagagactttgaatccaaagcaaacat 1935
Db 661 CACTCAGTGAGCACAGAGAGCGCTCATGAACGAAGAGACTTTGAATCCAAAGCAAACTAT 720
Qy 1936 cttgggtgattctgggtgagcctctggaagaccggagcattctctggaggcagcagata 1995
Db 721 CTTGGTGATCTGGTGGGACTCTGTGAAGACCCGGAGGCAATCTCTGGAGGCGAGATA 780
Qy 1996 ttctccagtgatgaccccccgcagagggcggtgctgattctccagcagatatgaagattat 2055
Db 781 TTCTCTCCGAGTAGCCACCCCGCAGAAAGCGGTGCGATTTCTTCCAGCAGATATGAAGATTAT 840
Qy 2056 tcagagctggagagcttcccccagatctctttagaaccagtttgaaagatggccc 2115
Db 841 TCAGAGCTGGGAGAGCTTCCCCACGATCTCTTTAGAACCAGTTTGTGAAGATGGCCCC 900
Qy 2116 ttggccccccaccagagaaagaaagacatctcgtgagctccgagagctgtggcaa 2175
Db 901 TTTGGCCCCCACCAGAGGAAAGAAAGAGACATCTGTFAGCTCCGAGAGCTGTGGCAA 960
Qy 2176 agggctattcttcaacagatactgctttagaatggagaagaaatacagaagctccaa 2235
Db 961 AAGGCTATTCTTCAACAGATACTGCTGTAGAAATGGAGAAGGAAATCAGAAGCTCCAA 1020
Qy 2236 gctctgaaatgattctgtaacagcctgagctgattatgaagaatactccc 2295
Db 1021 GCCTCTGAANAATGATTGCTGAACAGCCCTGAGCTCGATATGAGAANAATTAATCTCC 1080
Qy 2296 tgccttaagaagaactacagtgtgggaaagatgcttagcactccaggaagatacaaa 2355
Db 1081 TGCTTTAAAGAACTACTACAGTGTGGGAAAGATGCTTTAGCACATCCAGGAAAGATCAAAA 1140
Qy 2356 attaatgtgacatgaaataatgactcggctgttgggcaagggtgtgccagtcatac 2415
Db 1141 ATTAAGTTTGACATGGAAAAATGCACTCGGCTGTGGGCAAGGTGTGCCACGTATCAC 1200
Qy 2416 cgaggtgaaatctggaaattctagctgagcaattccacttaaacaccagtttcccagc 2475
Db 1201 CGAGGTGAATCTGGAAATTTCTAGCTGAGCAATTCACCTTAACACCCAGTTTCCAGC 1260
Qy 2476 aaacagcagccaaagatgtgcatacaagaactttaaagcagctgacttcccagcag 2535
Db 1261 AAACAGCAGCCAAAGGATGTGCATACAAAGAACTCTTAAGCAGCTGACTTCCCAGCAG 1320
Qy 2536 catgcatcttattgacattggggaaccttctctacacaccatacttctctgcccag 2595
Db 1321 CATGCGATTCTTATGACCTTGGGCAAGCTTTCTCTACACACCCATCTCTGCGCCAG 1380
Qy 2596 cttggagcagcagctatcgctttacaacatttgaagcctactcactcttagaccag 2655
Db 1381 CTTGGAGCAGGACAGCTATCGCTTTTACAATTTTGAAGGCTTACTTCTTAGACCAG 1440
Qy 2656 gaagtggatttgcgaagctcagctttagcagggcattttgcttctcatatgagt 2715
Db 1441 GAAGTGGGATATGCCAAGCTCTCAGCTTGTAGCAGGCAATTTGCTTCTCATATGAGT 1500
Qy 2716 gagggaagcgcttataaatgctcaagtttctgagtgttgacatggggtgcggaaacag 2775
Db 1501 GAGGAAGAGCGCTTTAAATGCTCAAGTTTCTGATGTTTACATGGGGCTGGGGAACAG 1560

Qy 2776 tatcgccagacatgattattttacagatccagatgtaccagctctcaggttgcattat 2835
Db 1561 TATCGCCAGACATGATTATTATTCAGATCCAGATGTACCAGCTCTCGAGGTTGCTTAT 1620
Qy 2836 gattaccacagagacctctacaatcacctcgaggagcagagatcggccccagcctctac 2895
Db 1621 GATTACCACAGAGACCTCTACAACTCACCTGGAGGACGAGATCGGCCCCAGCTCTAC 1680
Qy 2896 gctgccccctggttctcacaatgtttgcctcaacagttccccctgggattctagccaga 2955
Db 1681 GCTGCCCTCGTGTCTCACCATGTTGCTCACAAGTTCCCGCTGGGATTCGTAGCCAGA 1740
Qy 2956 gctcttgatagattttcttcaggaacagaggtcatatttaaagtggtcttaagtctg 3015
Db 1741 GTCCTTGTATGATGATTTTCTTCAGGGAACAGAGCTCATATTTAAAGTGGCTTTAAGTCG 1800
Qy 3016 ttgggaagccataaagcccttattctgcagcatgaaaacctagaaccatagttgacttt 3075
Db 1801 TTGGGAAGCATAAGCCCTTGATTCTGCAGCATGAAAACCTAGAAACCATAGTTGACTTT 1860
Qy 3076 ataaaaagcagctaccacaccccttggtgtgacagatggaaagacccatcaatcaggtta 3135
Db 1861 ATAAAAAGCACGCTACCCAACTTGGCTTGGTACAGATGGAAAAGACCATCAATCAGGTA 1920
Qy 3136 ttgaaatggacatcgctaaacagttacaagcttatgaagttgagtcaccacgtccctcaa 3195
Db 1921 TTTGAANTGGACATCGCTAAACAGTTACAAGCTTATGAAGTTGAGTACCACCTCTTCAA 1980
Qy 3196 gaagaacttatgattctctctctcagtgacacccaaagaatggataaattagagaaa 3255
Db 1981 GAAGAATTTATCGATTCTCTCTCTCAGTGACAAACCAAGAAATGGATAAATTAGAGAAA 2040
Qy 3256 accaagcagcagctacgaaacacagaaaccttgacctcttccttgaaacagttgcagtgccaaat 3315
Db 2041 ACCAACAGCAGCTTACGCAACAGAACCTTGACCTTCTTGAACAGTTGCGAGTGGCAAT 2100
Qy 3316 gtagagttcaagccttgagccaccattgagaagctcctgagcagtgagagcaagctg 3375
Db 2101 GGTAGGATCCAAAGCTTTGAGGCCACCATTTGAGAAAGCTCTCTGAGCAGTGAGAGCAAGCTG 2160
Qy 3376 aagcagggccatgcttaccttagaaactgagcgtggtggccctctcagcagcgtgagagag 3435
Db 2161 AAGCAGGCCATGCTTACCTTGAAGTGGAGCGGTGGCCCTGCTCGACAGCGTGGAGGAG 2220
Qy 3436 ctgcggcgagcagcagagcccgagcagcgtgagcgtgagtcgacgcagcccgagccc 3495
Db 2221 CTGCGGCGCGAGCGCAGAGCCCGAGCCGAGCTGTAGTCACGCGAGCCCGAGCCC 2280
Qy 3496 acggcgactgacagctctgcaggagagattgcaaacaccatccccactgtccaggcctt 3555
Db 2281 ACGGGCGACTGACAGCTCTGCAGGAGAGATTGCAACACCATCCACACTGTCCAGGCCCTT 2340

RESULT 5

AK027355
LOCUS Homo sapiens cDNA FLJ14449 fis, clone HEMBB1001684.
DEFINITION
ACCESSION AK027355
VERSION AK027355.1 GI:14041975
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens embryo, 10 weeks whole embryo, mainly body cDNA to mRNA, clone lib:HEMBB1 clone:HEMBB1001684.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (sites)
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H., Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Nakamura, Y., Nagahari, K., Masuho, Y., Minomiya, K. and Iwayanagi, T.
 NEDO human cDNA sequencing project
 Unpublished
 2 (bases 1 to 2362)
 Isogai, T. and Otsuki, T.
 Direct Submission
 Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
 Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3951, Fax: 81-438-52-3952)
 NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing;
 Research Association for Biotechnology; cDNA library construction,
 5'- & 3'-end one pass sequencing and clone selection; Helix
 Research Institute (supported by Japan Key Technology Center etc.)
 and Department of Virology, Institute of Medical Science,
 University of Tokyo

FEATURES

Location/Qualifiers

source

1..2362
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HEMB1001684"
 /tissue_type="whole embryo, mainly body"
 /clone_lib="HEMB1"
 /dev_stage="embryo, 10 weeks"
 /note="cloning vector: pME18SFL3"
 40..2064
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAB5057.1"
 /db_xref="GI:1401976"
 /translation="MLKNKARSLTESILSRGNKARGLOEHSIVDLSSLSSTL
 SNTSPSPVCEALPISSEPSFKLLGSSDLSDESHPPEAPLSPQAPRRANT
 LSHFPICEOPQPPARGSPQSKLMRYHSVSTETPHERKDFESKANHLGSGTPV
 KTRHSRQOIFLRVATPQKACDSSRYEDISELGELPPSPLEPCEDGPPGPPEE
 KRTSRELRLAKAILOQILLLRMEKENOKLOASENDLLRLKLDYBEITPCKEV
 TTVWEKMLSPGRSKITDKFMKHAAGVGVPRHRRGEIWKFLAQFLHKHFPSPKQV
 PKDVPYKELLGTLTQSQAAILDLGRTPHPYFSAQLGAGOLSLNILKAYSLDDOE
 VYVCGSLFVAGILLHMSFEAPFKMLFDMGLRKQYRDMILLOIQYOLSRLL
 HDYHRDLYNHLSEHEIGPSLYAAPFLFPASQFPLGFVARFVDFMIFLOGTEVFKYA
 LSLGSHKPLLILOHENLETIVDFIKSTLNPGLGVQMEKTINGVFMENIAKQLOAYEVE
 YHVLQELSDLSPLSDNRMDKLENTSKONLDLLEQLQVANGRTSLRATIEKL
 LSESKLKQAMLTLELERSALLQTVLEELRRRSAPERSDREPECTQPEPTGD"

CDS

/note="unnamed protein product"

1804 atcgatgcccaggaacctccacaacctgcccgggggtccccgggggttttcgcaaaagaaa 1863
 |||||
 361 ATCGAATGCCAGGAACCTCCACAACCTGCCGGGGTCCCGGGGGTTTCGCAAGAGAAA 420
 |||||
 1864 cttatgaggtatcactcagtgagcacagacgcctcatgaacgaagagactttgaatcc 1923
 |||||
 421 CTTATGAGGTATCATTACATCAGTGACACAGACGCCCTCATGACGAAGAGACTTTGAATCC 480
 |||||
 1924 aaagcaaacacatttgggtgattcttgggtggaactccttgaagaccggggagcttctg 1983
 |||||
 481 AAAGCAAAACCATCTTGGTGATTCTGGTGGGACTCCTGTGAAGACCCGGAGGCATCTCTGG 540
 |||||
 1984 aggcacagatatcttcgagtagtagccaccgccgcaagaaggcgtgcgattcttccagcaga 2043
 |||||
 541 AGGCACAGATATTCCTCCGAGTAGCCACCCCGCAAGAGCGTGCATTCTCCAGCAGA 600
 |||||
 2044 tatgaagattattcagagctgggagagcttccccccacgattcctcttagaacagattt 2103
 |||||
 601 TATGAAGATTATTACAGAGCTGGGAGAGCTTCCCCACGATCTCCTTTAGAACCAGTTTGT 660
 |||||
 2104 gaagatggggcccttggccccccaccagaggaagaaagagacatctctgagctccga 2163
 |||||
 661 GAAGATGGGCCCTTTGGCCCCCACCAGAGAGAAAGAAAGAGACATCTCTGAGTCCGA 720
 |||||
 2164 gagctgtgcaaaagctattctcaacagatactctgcttagaattgagagagaaat 2223
 |||||
 721 GAGCTGTGCAAAAGGCTATTCTTCAACAGATACTGCTGTAGATGGAGAGGAAAT 780
 |||||
 2224 cagaagctccaaagcctctgaaatgattgtgaaacaagcgctgaagctcgattatgaa 2283
 |||||
 781 CAGAAGCTCAAGCCTCTGAAATGATTGTCTGAACAAGCGCTGAAGCTCGATTATGAA 840
 |||||
 2284 gaattactcctctgtttaaagaagtaactacagtggtgggaaagatgcttagcactcca 2343
 |||||
 841 GAAATTAATCTCTGTCTTAAAGAAAGTAACATACAGTGTGGGAAAAGATGCTTTAGCATTCCA 900
 |||||
 2344 gaaagatcaaaaattgaattgacatggaaaataacacacacgctgttggcgaagtg 2403
 |||||
 901 GGAAGATCAAAATTAAGTTTGACATGGAAAAAATGCATCTGGGCTGTTGGGCAAGGTG 960
 |||||
 2404 ccacgtcatccagaggtgaaatctggaatttctagctgagcaatttccacccttaaacac 2463
 |||||
 961 CCAGCTCATCACCGAGGTGAAATCTGAAATTTCTAGCTGAGCAATTCACCTTAAACAC 1020
 |||||
 2464 cagttccacagaacacagcagcgaaggatgtgccatacaaaagactcttaagcagctg 2523
 |||||
 1021 CAGTTTCCAGCAACAGCAGCAGCAAGAGATGTGCCATACAAAGAACTCTTTAAAGCAGCTG 1080
 |||||
 2524 actccacagcagcatcgattcttattgacctggggcgaaccttctcacaacccatcac 2583
 |||||
 1081 ACTTCCAGCAGCATCGGATCTTATTGACCTTGGGGAACCTTTCTACACACCATAC 1140
 |||||
 2584 ttctctgcccagcttggagcagcagcagctatcgctttacaacattttgaaggcctactca 2643
 |||||
 1141 TTCTCTGCCAGCTTGGAGCAGCAGCAGTATCGCTTTTACAACTTTTGAAGGCTCTACTCA 1200
 |||||
 2644 ctctagacagagaagtggatattcccaaggtctcagctttgtagcagcattttgctt 2703
 |||||
 1201 CTTCTAGACAGGAAGTGGGATATTGCCAAGGCTCAGCTTTGTAGCAGGCATTTTCTT 1260
 |||||
 2704 ctctcatgagtgaggaaggcgtttaaaatgctcaagtctcagattgttgcacatgggg 2763
 |||||
 1261 CTTTCATATGAGTCAGGAAGAGCGCTTTAAATGCTCAAGTTTCTGTGATGTTGACATGGG 1320
 |||||
 2764 ctgcgaacacagtatcgccgacagacatgattattttacagatccagatgtaccagctctcg 2823
 |||||
 1321 CTGCGAAACAGATATCGGCCAGACATGATTTATTTTACAGATCCAGATGTACCAGCTCTCG 1380
 |||||
 2824 aggttcttcatgattaccacagacagccttcaaatcaacctggagagcagcagatcgcc 2883
 |||||
 1381 AGGTTGCTTCATGATTACCACAGAGACCTCTACAATCACTCTGGAGGACGACGAGATCGGC 1440
 |||||

BASE COUNT

657 a 581 c 578 g 546 t

ORIGIN

Query Match 59.4%; Score 2112; DB 9; Length 2362;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1444 agtgaattaccaccagtcgactgatttagctagatgctgaaaaaagcaag 1503

DB 1 AGTGAATTACCCAGTCGACTCGATTAGCTAGATGCTGAAAAACAAGCAAG 60

QY 1504 agatctttaacagagctttagaagattttgtccggggtaataaagcagagcctg 1563

DB 61 AGATCTTTAACAGAGTCTTTAGAAAGATTTTGTCCGGGGTAATAAAGCCAGAGGCCTG 120

QY 1564 caggaaacatccatcagctggtatcgatagctccctgcttagtaccattaaataacacc 1623

DB 121 CAGGAACACTCCATCAGTGTGGATCTGGATAGCTCCCTGTCTAGTACATTAAAGTAACACC 180

QY 1624 agcaaaagaccatctgtgtgaaaaggagcccttgcctcatctcagagctcctttaag 1683

DB 181 AGCAAAAGGCATCTGTGTGTGAAAAGAGGAGCCCTTGCCCCATCTGAGAGCTCCTTTAAG 240

QY 1684 ctctcggctcctcgaggagcctgtccagtgactcgagagtagtcatctcccgagaagagcca 1743

DB 241 CTCCTCGGCTCCTCGGAGGACCTGTCCAGTGACTCGGAGAGTCACTCCCGAGAGAGCCA 300

QY 1744 gctccgctgtgccccagcaggccttcaggaggcgagcaaacacctgacttccctccccc 1803

DB 301 GCTCCGCTGTGCGCCAGCAGGCGCTTCAGGAGGCGAGCAAAACACCTCTGAGTCACTTCCCC 360

Qy 2884 cccagcctctacgctgccccctgggttcctcaccatgtttgctcacagtttcccgctggga 2943
|||||
Db 1441 CCCAGCCTCTAGCTGCCCCCTGGTTCCTACCAATGTTTGCCTCACAGTTCCCGCTGGGA 1500
Qy 2944 ttctagccagagcttttgatgatatttttcttcagggaacagaggtcatattaaagt 3003
|||||
Db 1501 TTCTGAGCAGAGCTTTTGATATGATTTTCTTTCAGGGAACAGAGGTCTATATTTAAAGTG 1560
Qy 3004 gctttaagctctgtgggaagccataaagcccttgattctgcagcatgaaacctagaacc 3063
|||||
Db 1561 GCTTTAAGCTCTGTGGGAAGCCATTAAGCCCTTGATCTGCGACATGAAGAACTAGAAACC 1620
Qy 3064 atagtgtactttataaaaagcaagcctaccccaaccttgcttggttacagatggaagaagacc 3123
|||||
Db 1621 ATAGTTGACTTTATAAAAAGCAGGCTACCCAACTTGGCTTGGTACAGATGGAAGAGACC 1680
Qy 3124 atcaatcaggtattgaaatggagcaatcctgaacagcttcaagcttatgaagtgaagtac 3183
|||||
Db 1681 ATCAATCAGGTATTTGAAATGGACATCGCTAAACAGTTTACAAGCTTATGAAGTTGAGTAC 1740
Qy 3184 cagctcctcaagaagaacttatcgattcctctcctcctcctcagtgacacaaagaagtgat 3243
|||||
Db 1741 CAGTCTCTCAAGAAGAAGCTTATCGATTCCTCTCTCAGTGACCAACCAAGAAATGAT 1800
Qy 3244 aaattagagaaaacacacagcagcttaccgcaaacagaaaccttgacctctcttgaaacgttg 3303
|||||
Db 1801 AAATTAGAGAAAACCAACAGCAGCTTACGCAACAGAACTTGACCTCTCTGAACAGTTG 1860
Qy 3304 caggtggcaaatgtagatccaaagccttgaggccaccattgagaagctcctcagcagt 3363
|||||
Db 1861 CAGGTGGCAATGTGATGATCCAAAGCCTTGAGGCGCACCATTTGAGAAGCTCTGTAGCAGT 1920
Qy 3364 cagagcaagctgaagcagcagccttacccttagaactgagcgtcgccctcgtcag 3423
|||||
Db 1921 GAGACAGCTGAAGCAGGCCATGCTTACCTTAGAAGCTTGAGCGGTGGCCCTGCTGCGAG 1980
Qy 3424 acggtggaggagctgcgagcgagcagcagccagcagcagcgagcctgagtgacag 3483
|||||
Db 1981 ACGGTGGAGGAGCTCGCGCGGAGCGAGCGCAGAGCCAGCGGAGCCTGAGTGACG 2040
Qy 3484 cagcccgagccacagggcgagctgacagctctgcagagagagattgcaacacatccacac 3543
|||||
Db 2041 CAGCCCGAGCCACAGCGGAGCTGACAGCTCTGACAGAGAGATTGCAACACCATCCACAC 2100
Qy 3544 tctcaggcctt 3555
|||||
Db 2101 TGTCAGGCGCTT 2112
|||||

RESULT 6
BTAL7923
LOCUS BTAL7923 1995 bp mRNA linear MAM 18-NOV-1998
DEFINITION Bos taurus mRNA for linceln.
ACCESSION Y17923
VERSION Y17923.1 GI:3893076
KEYWORDS linceln; linceln.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 1995)
Xu, X. Z., Wes, P. D., Chen, H., Li, H. S., Yu, M., Morgan, S., Liu, Y. and
Montell, C.
Retinal targets for calmodulin include proteins implicated in
synaptic transmission
J. Biol. Chem. 273 (47), 31297-31307 (1998)
99030403
2 (bases 1 to 1995)
Montell, C.
Direct Submission
Submitted (11-AUG-1998) C. Montell, The Johns Hopkins University
School of Medicine, Department of Biological Chemistry, 725 N.

Wolfe Street, Baltimore, MD 21205-2185, USA
Location/Qualifiers
1. .1995
/organism="Bos taurus"
/db_xref="taxon:9913"
/tissue_type="retinal"
/dev_stage="adult"
2. .1672
/gene="lyn"
CDS
/gene="lyn"
/codon_start=1
/product="linceln"
/protein_id="CAA76943.1"
/db_xref="GI:3893077"
/translation="ARAGVSQRKLVYHVSIVSTETPHERKDPESKADHISASRTPVKT
RRHSWRQIIFLRVATPKACSPKRYEDYSELGELPPSPLEPVCEGPGFVPEEK
RTSHELRLWKQAILQIILLRMKENGKQVSRHHRGEIWKFLAEQYHLKAYLHPCKQOPK
WKEILSTPKRSKIIFDMERKMSAVGGVQPHRRGEIWKFLAEQYHLKAYLHPCKQOPK
DTYKELKQITSSQHAIIIDLGTFTPHYYSALGAGQSLYNIILKAYSLLOEVG
YCQGLSFVAGILLHMGEEBAFNMLKFLMDGMRKQYRPDMIILOQMYQLSLLHD
YHRLYNHLEHEHEIGPSLYAAPWFLTFASQFPLGFVARVDFMIFLOGSEVIFKVALS
VLGSHKPLIIOHENLEIYDFIKSTPLNLGLVQMEKTSIQVFETDLSKQLOAYEVEYH
LLQBELIDSSPLSDNQMDKLEKTNLSLRKQNLLOLLOLQVANGRIOSLEATVEKLLT
SESKLKQATLALERALSALLOTVEQLRRQTAEIQLGQSDPTLPKPSGD"
BASE COUNT 560 a 516 c 497 g 422 t
ORIGIN
Query Match 38.2%; Score 1359.2; DB 4; Length 1995;
Best Local Similarity 88.4%; Pred. No. 1.9e-286;
Matches 1475; Conservative 0; Mismatches 193; Indels 0; Gaps 0;
Qy 1844 cgggggttcacaaagaaacttatgaggtatcactcagtcagtcagcagcagcctcatg 1903
Db 9 CAGGGGTCTCGCAAGAGAACTTGTGAGGTATCATCTAGTGAGCACAGAGCCCTCATG 68
Qy 1904 aacgaaagacttgaatccaaagcaaacactcttggtgattctctggtgagcctctga 1963
Db 69 AACGAAAGGACTTGTAGTCCAAAGCCGACCATCAGCGATGCCAGGAGGCCCTGTGA 128
Qy 1964 agaccggagcattcctcggagcgagcgatattctcctcgagtagccaccgcgagaagg 2023
Db 129 AGACACGGAGACACTCGTGGAGGCGACAGATCTTCTCTCGGTGGCCACCCCGCAGAAG 188
Qy 2024 cgtgcgattctccagcagatataaagattattcagcagcagcagcctcccccacgat 2083
Db 189 CCTGGAGTCCCCAAAAGATACGAAGATTACTCAGAGTTGGGAGAGCTTCCCCCACCAT 248
Qy 2084 ctcttttagaaccagtttgtgaagatgggcccctttggccccccaccagaggagaaagaaa 2143
Db 249 CTCCACTAGAACCACTTGTGAAGATGCACCTTTTGGCCGGTCCCGAGAGGAAAGAAA 308
Qy 2144 ggacatctcgtgagctcgcgagcgtgtggcaaaagggctattcttcaacagatactgtgc 2203
Db 309 GAACATCTCATGAACCTCGGGAGCTGTGGCAAAAGGCCATTCTACACAGATATTGCTGC 368
Qy 2204 ttgaatgagaagaaaaatcagaagctcccaagcctctgaaatgatctgtggaacaagc 2263
Db 369 TCAGATGGAGAAAGAAACCAAAACCTGCAAGCCCTCTGAGNATGATTTGCTGACACAGC 428
Qy 2264 gcctgaagctcgattatgaagaaattactcctcttcttaagaagaactacagctgtggg 2323
Db 429 GCCTAAAGCTTGATTATGAAGAAATCACTCCCTGTCTTAAAGAAAGTACTACAGTGTGG 488
Qy 2324 aaagatgcttagcactccaggaagatcaaaaattaaatttgacatgaaaaaatgcact 2383
Db 489 AAAAGATATTAGCAGCTCCAGGAAGATCAAAATTAAGTTTGACATGGAAGAAATGCAC 548
Qy 2384 cggctgttggcgaaggtgtgcacgtcatcccgaggtgaaatctctggaaatttctagt 2443
Db 549 CGGCTGTGGCCAGGTGTGCGCGACACACCACCGGGGTGAATCTGGAATCTTGTGGCGG 608

Db 61 TCCCTTTAGAACCGGTGTGTAGGACGGCCATTTGGCCCATACAGGAGAAAGAGG 120
QY 2143 agacatctctgtagctcogagagctgtgcaaaaagcttatttccaacagatactctgtg 2202
Db 121 AAGACGCTCAGCGGAGCTTCGAGAGCTGTGGAAGAGGCCATCTTGACGAGATCCTGCTG 180
QY 2203 ctLagaatgagagaaatcagaagctccaagctctgaaaatgatttctgtaacaag 2262
Db 181 CTCAGATGGAGAGGAGAAATCAGAGCTTACAAGCCTCTGAAAACGCTTTGCTGGAACAA 240
QY 2263 cgctgaagctgatttatgaagaaattactccctgtctttaaagaagtaactacagtggtg 2322
Db 241 CGCCTCAAGCTTGACATGATGAATCACTCCGCTGCTTAAAGAACTCACTACAGGTGG 300
QY 2323 gaaaagatgcttagcactccaggaagatcaaaaatttaagtgtgacatggaaaaaatgcac 2382
Db 301 GAAAAGATGCTTAGCACTCCAGAGATCCAAAATTAAGTTTGACATGGAAGAAAGTGCAC 360
QY 2383 tcgctgtgtggcaaggtgtgcacatcatccagaggtgaaatctgaaaatttctagct 2442
Db 361 TCAGCTGTGGGCAAGGTGTGCCAGCTCATCCAGAGGTGAGATCTGGAAATTTCTAGCT 420
QY 2443 gagcaattccacccttaaacaccagtttccagcaaacagcagccaaaggtgtgcatac 2502
Db 421 GAGCAGTTCACCTTAACACACCCATTTCTAGTAACACAGCGCAAGGACGTGCCCTAC 480
QY 2503 aagaactcttaagacagctgacttcccagcagcatgcgatttcttattgaccttgggca 2562
Db 481 AAAGAGCTCTGAAGAGCTGACCTCGCAGCAGCGCCATTTCTATCGACCTCGGGCGA 540
QY 2563 accttccctacacacactactctctgccagcttggagcagcagctatcgctttac 2622
Db 541 ACCTTTCCACACATCCATACTTCTCTGCCAGCTTGGAGCAGGTGACGTGTCACATTTAC 600
QY 2623 aacatttgaagcctactactcttagaccaggaagtggtgatatggcaaggtctcagc 2682
Db 601 AACATTCTGAAGCCCTACTCGCTCTGGACACGAGGTTGGATCTGCCAAGTCTCAGC 660
QY 2683 ttgtgacaggaattttgtcttcttcataagtgaggaagagcgttttaaaatgctcaag 2742
Db 661 TTTGTGGCAGGCATTTTGTCTTCATGATGAGGAGAGAGCGGTTCAGATGTCACAG 720
QY 2743 ttctgatgtttgacatgggctgcggaacagatctcggcagacacatgattattttacag 2802
Db 721 TTCTGTATGTTGACATGGGCTCGCGAACAATATCGCCACACATGATTTTTCAG 780
QY 2803 atccagatgtaccagctctcogaggtgtcttcattgattaccacagagacctctacaatcac 2862
Db 781 ATCCAGATGTACCAGCTGTACCGGCTCTCCACGATTACCCGAGACCTCTACAACCCAC 840
QY 2863 ctgagagacagagatcgcccccagcctctacgtcgccccctgttctcaccatgttt 2922
Db 841 CTGGAAGACAGCAGATCGGCCCGCCCTCTAGCGGGCTCCCTGGTTTCTCACCGGTTC 900
QY 2923 gcttcacagtcccgctgggattcgtagcagagcttttgatgatgtattttcttcaggga 2982
Db 901 GCCTCAGATTCCTCCATCGGCTTTGTAGCCAGAGTCTTTGATATGATCTCTCTCAGGA 960
QY 2983 acgaggtcatatttaagtgggttttaagtctgttggaagccataagcccttgattctg 3042
Db 961 TCAGAGGTCTATATTTAAAGTAGCTTTAAAGTCTTTTGGGAGCCATAAGCCCTTGATCTA 1020
QY 3043 cagcatgaaacactagaacactagttgactttataaaagacagctaccacacttggc 3102
Db 1021 CAGCATGAGAACCTGGAAACCATCGTGGACCTTCATAAAGAACACACTCCCCCAACCTGGGC 1080
QY 3103 ttggtcagatggaagaccatcaatcaggtatttgaatggacatcgctaaacagtta 3162
Db 1081 CTGGTCGATGAGGAGAACCATCAGTCAGGTGTTTGAGATGGACATCGCCACAGAGCTC 1140
QY 3163 caagcttatgaagtttagtaccacgtctcttcaagaagaacttatcgattctctctctc 3222

Db 1141 CAGCCCTATGAGGTGAGTACCACGTGCTCCAGGAGGAGCTTATTGAGTCTCGCCTCTC 1200
QY 3223 agtgacaacaaagaatgataataattagagaaaacaaacagcagcttagcagaacaagaac 3282
Db 1201 AGTGACAACCAAGAATGGAGAAATTTGGAAACCAACACAGCAGCTTGGCCAAACAGAAC 1260
QY 3283 cttgacctcttgaaacagttgcaggtggcaaatggtagtatccaaagccttggagccacc 3342
Db 1261 CTGAGCTCTCTGGAGCAGTTGCAGGTGGCAATGCTAGGATCCAAAGCCTTGAAGCCACG 1320
QY 3343 attgagaactctctgagcagctgagagcaagctgaaacagcagccatcttacttagaacty 3402
Db 1321 GTAGAGAAACTTCTTACCAGCGAGAGTAAGCTGAAGCAGGCTGCGCTGACCTGGAGTG 1380
QY 3403 gagcgtcgccctgtgcagacggtggagagagctcgccgcgggcagcagagccaccagc 3462
Db 1381 GAGCGTCCGCCCTGTCTGCAGATGTTGGAGGAGCTCGGAGGCAAGCCCGGCCCGCAGC 1440
QY 3463 gaccggagcctgagtcgacgagcccgagcccgccacggcgagcagctgacagct 3512
Db 1441 ACTCCAGAGCCAGCTGCACCCAGCTGGAGCCACAGCGGATTGACCGCT 1490

RESULT 8

AK057182 1788 bp mRNA linear PRI 31-OCT-2001
LOCUS Homo sapiens cDNA FLJ32620 fis, clone STOMA2000386, highly similar
DEFINITION to Bos taurus mRNA for lycein.
AK057182
ACCESSION AK057182.1 GI:16552778
VERSION oligo capping; fis (full insert sequence).
KEYWORDS Homo sapiens stomach cDNA to mRNA, clone_lib:STOMA2
SOURCE clone:STOMA2000386.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (sites)
AUTHORS Suzuki,O., Sasaki,N., Aotsuka,S., Shoji,T., Ichihara,T.,
Shiohata,N., Matsumoto,K., Hirano,M., Sano,S., Nomura,R.,
Yoshikawa,Y., Matsumura,Y., Moriya,S., Chiba,E., Momiyama,H.,
Onogawa,S., Kaeriyama,S., Satoh,N., Matsunawa,H., Takahashi,E.,
Kataoka,R., Kuga,N., Kuroda,A., Satoh,I., Kamata,K., Takami,S.,
Terashima,Y., Watanabe,M., Sugiyama,T., Irie,R., Otsuki,T.,
Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fuji,A.,
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project

TITLE

Unpublished
2 (bases 1 to 1788)
Isogai,T., Otsuki,T. and Sugiyama,T.

REFERENCE

Direct Submission
Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)

AUTHORS

TITLE
JOURNAL
COMMENT
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: RAB and
HRI.

FEATURES

source

Location/Qualifiers
1..1788
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="STOMA2000386"
/tissue_type="stomach"
/clone_lib="STOMA2"
/note="cloning vector: pME18SFL3"

BASE COUNT ORIGIN	481 a	412 c	434 g	461 t
Query Match Best Local Similarity Matches 999; Conservative	28.1%; 99.9%;	Score 998.4; Pred. No. 1.3e-207;	DB 9; Mismatches 1;	Length 1788; Indels 0; Gaps 0;
Qy	2556	tggcggaaccttctctacacaccatacttctctgcccagcttgagcagagacagctatc	2615	
Db	553	TGGCGGAACCTTCTCTACACACCCTACTTCTCTGCCAGCTTGAGCAGGACGACTATC	612	
Qy	2616	gctttacaacatttgaaagcctactcactctctagacaggaactggatattccaaagg	2675	
Db	613	GCTTTACAACTTTTGAAAGGCTTACTCTTCTAGACAGGAAGTGGATATTCCTCAAGG	672	
Qy	2676	tctcagcttttagcagcattttgctctcttcatactagtgagagaagcggttttaaaat	2735	
Db	673	TCTCAGCTTTTAGCAGCAATTTTGTCTTCTCATATGATGAGTGAAGAGGCGTTTAAAT	732	
Qy	2736	gctcaagttctgatgttgatcattggtgctcggaacacagtatcggccagacatgattat	2795	
Db	733	GCTCAAGTTCTTGATGTGACATGGGCTCGGAAACAGTATCGCCAGACATGATTAT	792	
Qy	2796	ttcacagatccagatgtaccagctctcgaggttcttcattcatgtattaccacagagacctta	2855	
Db	793	TTTACAGATCCAGATGATACCACTCTCTCGAGGTTCCTCATGATTTACACAGAGACCTTA	852	
Qy	2856	caataccttgagagcagcagatcggcccccagcctctacgctgcccctgggttcctcac	2915	
Db	853	CAATCACCTGGAGGACGACGAGATCGGCCCCAGCCTCTACGCTGCCCTGTTCTCCAC	912	
Qy	2916	catgtttcctcaagttcccgctgggattcgttagcagagctcttgatgattttct	2975	
Db	913	CATGTTTCCCTCACAGTTCCTCGGCTGGGATTCGTAGCCAGAGCTCTTGATATGATTTTCT	972	
Qy	2976	tcagggaacagagctcattattaaagtggctttaaagctgttggaagccataagccctt	3035	
Db	973	TCAGGGAACAGAGGTCATATTTAAAGTGCGCTTAAAGTCTGTGTGGAGCCATAGACCTT	1032	
Qy	3036	gattctgcagcatgaaacacctagaaaccatagttgactttataaaagcacgctaccocaa	3095	
Db	1033	GATTCTGCAGCATGAAACCTTAGAAACCATAGTTGACTTTATAAAAGCAGCTACCCAA	1092	
Qy	3096	ccttgcttgtagcagatggaaagaccatcaatcaggattttgaaatggacatcgctaa	3155	
Db	1093	CCTTGGCTTGTACAGATGGAAGAACCATCAATCAGGTATTTGAAATGGACATCGCTAA	1152	
Qy	3156	acagttacaagcttatgaagttgagtagcagctctcctcaagaagaacttatcgattcctc	3215	
Db	1153	ACAGTTACAAGCTTATGAAGTTGATGATGATGATGATGATGATGATGATGATGATGATG	1212	
Qy	3216	tcctctcagtgacacacaaagatggataatagagaaaaccacacagcagcttacgcaa	3275	
Db	1213	TCTCTCAGTGACAAACCAAGATGGATAAATTAGAAGAACCAACACAGCAGCTTACGCAA	1272	
Qy	3276	acagaaccttgacctccttgtagcaggttgaggttgaggttgaggttgaggttgaggttgag	3335	
Db	1273	ACAGAACCTTGACCTCTCTTGAACAGTTGACGTTGACGTTGACGTTGACGTTGACGTTGA	1332	
Qy	3336	ggccaccattgagaagctcctgagcagtgagagcaagctgaagcagggcctgcttacctt	3395	
Db	1333	GGCCACCAATTGAGAAGCTCTTGACGATGACAGCAAGCTGAAGCAGGCGCATGCTTACCTT	1392	
Qy	3396	agaactgagcgggtgagcctgctgacagcagctgagagagctgagcggcgagcagcaga	3455	
Db	1393	AGAACTGAGCGGTGCGGCCCTGCTGACAGCGGTGAGGAGCTGCGGCGCGGAGCGCAGA	1452	
Qy	3456	gcccagcagcgggagcgtgagtgacacagcccgagcccgagcggcgactgacagcgtctg	3515	
Db	1453	GCCCAAGCAGCGGAGCGCTGAGTGCACGACGCCCGAGCCACGCGGCGACTGACAGCTCTG	1512	
Qy	3516	caggagagattgcaac	3555	

Db 1513 CAGGAGAGATTGCAACACCATCCACACTGTCACAGGCTT 1552

RESULT 9
AB011175 5922 bp mRNA linear PRI 10-APR-1998
LOCUS Homo sapiens mRNA for KIAA0603 protein, complete cds.
DEFINITION
ACCESSION AB011175
VERSION AB011175.1 GI:3043729
KEYWORDS KIAA0603 protein.
SOURCE Homo sapiens male brain cDNA to mRNA, clone_lib:pBluescriptII SK plus clone:HG1488b.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5922)
AUTHORS Ohara,O., Nagase,T. and Ishikawa,K.
TITLE Direct Submission
JOURNAL Submitted (13-FEB-1998) Osamu Ohara, Kazusa DNA Research Institute, DNA Technology, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:cdmainfo@kazusa.or.jp, tel:+81-438-52-3913, Fax:+81-438-52-3914)
2 (sites)
REFERENCE
AUTHORS Nagase,T., Ishikawa,K., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes. IX. the complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro
JOURNAL DNA Res. 5 (1), 31-39 (1998)
MEDLINE 98290545
FEATURES
Location/Qualifiers
1..5922
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HG1488b"
/sex="male"
/tissue_type="brain"
/clone_lib="pBluescriptII SK plus"
348..4247
/gene="KIAA0603"
348..4247
/gene="KIAA0603"
/codon_start=1
/product="KIAA0603 protein"
/protein_id="BA025529.1"
/db_xref="GI:3043730"
/translation="MEPPSCIQDEPPHLEPFGVSAQPGPKPDKRRLRLWTVGSG
CLDRHTLPLMLMAEIRRSQKPEAGCGCAPAAREVILVLSAPFLURCPAPGAGAS
GGTSATQPNPAVFIEPKAHIISREIHNSHDLTYPAYLIKAPDDPESOMACHVFR
ATDPQVDPDISSIROLSKAAMKEDAKPKDNEDAFYNSQKFEVLYGKVTYVHKRAP
SSLIDCMKFSIHQORLKIQGEORGPDEGLADLEVVPVPGSGCLPEEADGTOT
HLGLPAGASQPALTSRVCFPERILEDGDEQEFRRSCSVTVGVORRHGSKSQ
PRRHASAPSHVQPSDEKRNITLFWGFEINLISDTKSVLEKFNKDISSCSQI
KHYDHFGICRESPEGLSQICYFQCASESLVDEYMLTKQAFSTAAALQSAKTQI
KLCAECPMHSILKICERIEGLYPRAKLVIOHLSLTDNEQADIFERVQKMPVSDO
EENELVTLHLRCLCAKQKTHVIGEPSTISNTIPENATSSGRFKDLILKNKAKRS
LTSLENIISFGANRMRGLSGVDSFERSNLSAKEDYSPGDSPGTPPASPPSSAWQ
TFDESDSPQFRRRAHTFHPSPSTKKNLQDGRAQGVRSPLLRGSSSEQCSNLS
VRRMYKESNSSSLPSLHTSFSAPSTFSLKSFYONSGRLSPQYENETKQDTASES
SDGEGKRTSTCSNESLVSQGVSTVPRRISWRQRIFLVASPNKSPKSAQQQDGLD
RNELLPLSLPSTMEEPVIFLGSDEDDPEIKERKSKELRSLMKRAKIRCDMDIHT
EKENKLEGASRDELOSKVKLDYEEVGAQKEVLIWDKLLKNCRAKIRCDMDIHT
LLEGVPKSRGEITWQFALQYLRHRLPNKQPPDISYKELLKQLTAQOHAIVLDIG
RTPTPTFYVQLGPGQLSNLLKAYSLDLKVEYGCQIGISFVAGVLLHMSSEQAE
MKFLMYDLGFRKOYRDMMSLIQOMYQLSKLLHDYHRDLNHLNEEISFSLYAAPW
MLFLAFQSFLGVARVDIIFLQGTVEVIFKVALSLSSQETLMECESENVIEFLK
NLTLDPMNTSEMKIITOVFEMDISKOLHAYEVEYHVLQDELQESSYCESEETLEKLE
RANSOLKRONMDLLEKLOVAHTKIOALESNLENLLTRETMMKSLIRTLEQEKMAQKT
VEOLRKLPPADALANCLDLLRLDNCNPNKAKIGNKP"
BASE COUNT 1624 a 1408 c 1432 g 1458 t
ORIGIN

Query Match		19.2%	Score 682.2:	DB 9:	Length 5922;
Best Local Similarity		56.6%	Pred. No. 1.9e-138;		
Matches 1585; Conservative		0;	Mismatches 1028;	Indels 186;	Gaps 10;
QY	823	agcggacacaaatattgtcagccacacagatatcgaggaaaatgaaactatgctcttcacg	882		
DB	1368	AGCGACCCAGTCAGGTCACGCTCCAGGCTCGGACTCGGAGAGAACAGGACCATGCTCTCCAG	1427		
QY	883	attggccagctcgaagtaccctaccatcagctcagacccaaaaaataagcattggagaaa	942		
DB	1428	GTGGGCGAATTCAGATTAACCTTATCAGTCCAGACACTAAATCAGTTGCTAGAAAAG	1487		
QY	943	aatttaagagagatatcctttgctctcagggcatcagacacgtgacacactttgggttt	1002		
DB	1488	AAATTTAAAGATATCTCTGTTCTCAGGGTATAAAGCATGTGGATCACTTTGGCTTT	1547		
QY	1003	atctgtgggagcttcctcggaggtggcggtttcattttgtctgtctacgtgtttcagtc	1062		
DB	1548	ATCTGCCGGGAGTCTCCAGAGCTGGACTTAGCCAGTATATTTGTTATGTATTCACGTGT	1607		
QY	1063	acaaatagagctctggttgatgataaattatgatgacctgaacaggcccttcacgggtggcc	1122		
DB	1608	GCCAGCGAATCTCTGGTGTATGAGTAAATCTGACTCTGAACAGCGCCTTCAGTACGGCG	1667		
QY	1123	gcagtcagcagacagcagtaagggcgccagcccgctgtgtgagggctggccctcgcaaacg	1182		
DB	1668	GCTGCCCTGCAGAGTGCCAGAGCGCAGATTAACCTGTGTGAGGCGCTGCCGATGCACTCT	1727		
QY	1183	ctgcacaagctctgtgagaggtagggggaatgaattcttccaaaaaacaactagaactg	1242		
DB	1728	TTGCATTAAGCTCTGTGAAGGATTTGAAGGTCTCTACCCACCAAGAGCAAGCTGGTGATA	1787		
QY	1243	caaaagcacctgacgacattaaaccaatcaggagcagcgactattttgaagggttcag	1302		
DB	1788	CAGAGGCATCTCATCATCCTGACAGATAAATGACAAAGCTGACATCTTTGAAGAGTTCAG	1847		
QY	1303	aaattggagccgagaagatgcagcgagagagaatgaattgtattttcttctgagatgt	1362		
DB	1848	AAATGAAGCCAGTCAGTGACCGAGGAAGAAATGAACCTTGATTTTACACCTTGAGGCAG	1907		
QY	1363	ttatatgaagagaaacagaagaaacacatccattattggggagatgaagcagacatgcag	1422		
DB	1908	CTGTGTGAAGCCAGCAAGAGACACAGTGCACATCGGGG-----AAGGCCCTTCT	1958		
QY	1423	atggcagcagaagaatattggaagtgaattaccaccagtgccactgattagctagat	1482		
DB	1959	ACTATTTCAATAGTACAATCCAGAAAATGCAACAGCAGTGGAGGTTCAACCTTGAC	2018		
QY	1483	atgctgaaaaacaaagcaagagatctttaacagagctctttagaaagtattttgtcccg	1542		
DB	2019	ATTCTGAAAANTAAAGCTAAGAGATCCTTAACCTAGCTCCCTGGANAATATCTTCTCAGG	2078		
QY	1543	gg---taataaagccagagccctgcaggaacactccactcagt-----	1581		
DB	2079	GGAGCTAACAGATGAGAGTGCGCTTGGAAAGTGTGGACAGTTTGAACCGTCCAACAGT	2138		
QY	1582	-----	1581		
DB	2139	CTTGCTTCAGAGAGGACTACTACAGGGGATTTCTCCACAGGGATTCACAGGCAGCGGTCC	2198		
QY	1582	-----gtgagctgtgtagctcctcctgtcagt	1608		
DB	2199	CCACCGCTCTCAGCTTGGCAACGTTTCCGGAAGAGGATTCGGACCTCCCGCAGTTTCGA	2258		
QY	1609	acattaagtaacaccagcaagggccatctgtgtgtgaaaaagaggccttgcacctctct	1668		
DB	2259	AGACGGGCACACAGGTCAGCCACCCACCTTCAGCAACAAAGAGAAAGCTGAATTTGCAG	2318		
QY	1669	gagagctcctttaagctcctcgtcctctcg-----gagg	1702		
DB	2319	GATGGGAGGGCTCAGGGTGTGCGTTCCCTCTGCTGAGGAGAGCTCCAGTGAACAGTGC	2378		

QY	1703	acctgtccagtgactcgagagtcactctcccaagaagccagctccgctgtgccccagc	1762		
DB	2379	AGCAATCTTTTCGTCAGTTCGACGCATGTACAGAGAGATTAATTTCTCTCCACGTCCTCCA	2438		
QY	1763	agccttcaggagcgagcaaacacacccctgagtcacttccccatcgaatgccaggaacctc	1822		
DB	2439	AGTCTTACACTTCTCTGCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	2498		
QY	1823	cacaaactgcccgggggtc---cccgggggttcgcgaaggaacacattatgaggtatcact	1879		
DB	2499	TACCAGAAATTCAGGTAGACTGTCCCCACAGTATGAAAATGAAATCAGACAAAGACACTGT	2558		
QY	1880	cagtgcagcagagacgcctcatgaacgaagagctttgtaattccaaagcaaacactcttg	1939		
DB	2559	TCAGATCAAGTGTATGGAGNAGGAGAAAAGACCTCATCTACCTGCACCAATGAGTCC	2618		
QY	1940	gtgattct-ggtgggactcgtgaagacccggaggacattcctcctggggagcgagatattc	1998		
DB	2619	CTAAGTGTGGAGAAACCTCTGTCTCTCTGCGCGATCTCTCTGGCGCAGCGCATTTTC	2678		
QY	1999	ctcggagtgcaccccgagaaagcggtcgattcttcagcagagatatgaagattattca	2058		
DB	2679	CTCAGGGTGTCTCTCCATGAACAAATCTCCCTCAGCAATGCAACAGCAAGATGGATTG	2738		
QY	2059	gagctgggagagcttccccacagatctccttttagaacacagtttgtgaagatgggcccctt-	2117		
DB	2739	GACAGNACGAGCTGCTGCCACTGTCCCCCTCTCTCCACCATGGAGGAGGAACCGCTG	2798		
QY	2118	-----tgccccccaccagagagaaaaagaaaggaacatctcgt	2154		
DB	2799	GTTTATATTCCTGTCTGGGAGGATGACCCAGAAAAGATTTGAAGAAAAGAAATCAAAA	2858		
QY	2155	gagctcggagagctgtggcaaaagctattcttcaacagatactgctgcttagaaggag	2214		
DB	2859	GAATCAGAGAGCTGTGGAGAAAAGCTATACACCACAAATCTTGTACTTCGAATGGAA	2918		
QY	2215	aaggaataacagagctcca---agcctctgaaaaatgatttgtgacaaagcgccctgaag	2271		
DB	2919	AAAGAAAACAGAACTTGAAGGAGCAAGCAGAGATGAACCTCCAGTCCACAGAAAGTTAA	2978		
QY	2272	ctcagattatgaagaaattactcctctttaaagaagtaactacagtggtggaaagatg	2331		
DB	2979	TTAGACTATGAAGAAAGTTGGTGTCTGCAAGAAAGGCTTTAATAACTTTGGGATAAAG	3038		
QY	2332	cttagcactccaggaagatcaaaaattagtttgcagtgaaagaaatgcactcgctgtt	2391		
DB	3039	TTGTTTAA---ACTGCAGAGCTAAATCAGATGTGATATGGAAGATATTCATCTCTCTT	3095		
QY	2392	gggcaaggtgtgccacgctcatcccgagggtgaaaatctggaaaattcttagctgggcaattc	2451		
DB	3096	AAAGAAAGGAGTTCCTCAAAAGTCAGCAGGAGAGAAATTTGGCAGTTTCTGGCTTTACAGTAC	3155		
QY	2452	cactttaaacacacagtttcccagcaaacagcagcaaaagatgtgcatacaagaactc	2511		
DB	3156	CGACTCAGACAGATGTGCTTAATAAACACAGCCCTCTGACATATCTCTATAAGAACTT	3215		
QY	2512	ttaaagcagctgacttcccagcagcagctcttcttattgaccttgggcaaccttccct	2571		
DB	3216	TTGAGCAGCTCACTGCTCAGCAGCATGCGATTCCTGTGGATTTAGGAAGGAGCTTCTCT	3275		
QY	2572	acacacccatctctctgcccagcttgagcagggagacagctatcgctttcacacatttg	2631		
DB	3276	ACTCACCTTACTTTTTCAGTACAGTGTGGCCAGGACAGCTGTCTACTGTTTAACTCCCTG	3335		
QY	2632	aagccttactcactcttagaccaggaagtgggatattgccaaagctccagctttgttagca	2691		
DB	3336	AAAGCTATTCTTTGCTGGCAAAAGAGTGGGATCTGTCTCAGGGGATCAGCTTTGGGCT	3395		
QY	2692	ggcatttgcctctcatatgagtgaggaagggcgttttaaaatgctcaagttctctgatg	2751		
DB	3396	GGAGTCTGCTTCTGCACATGAGTGAAGCAAGCGCTTTGAAATGCTGAAATTCCTCATG	3455		
QY	2752	tttgacatggggctcggaacacagtatcgccgagacatgattattttacagatcccgatg	2811		


```
Db 3456 TATGACCTCGCTTCGCAACGACATGACAGACCTGACATGATGTCGCTGCAGATTCAATG 3515
Qy 2812 taccagctctcaggttgccttcattgattacacagagacccctctacaatcacctggagag 2871
Db 3516 TACCAGCTGTCCAGGCTCCTTCATGACTATCACAGAGATCTCTACAATCACTACCTTGAAGAA 3575
Qy 2872 cagagatcgcccgccctcagctcagctgcctccctggcttcctcaccattgttgctccacag 2931
Db 3576 AATGAATACGCCCCAGCTTTATGCTGCCCCCGGTGCTTCACATGTTTGGCCCTCTCAG 3635
Qy 2932 ttccgcctggattcgtagccagagctcttgatgatgattttcttcagggaacagagctc 2991
Db 3636 TTTTCATTAGGATTGTAGCCAGAGTTTGTGATATTATTTCTTCAGGGAACCTGAAGTT 3695
Qy 2992 atattaaagtggcttaagtctctgtgggaagccataagcccttgatctcgcagcatgaa 3051
Db 3696 ATATTCAAGGTTGCACTCAGCCTACTGAGCAGCCCAAGAGACACTTAATAGGAATGTGAG 3755
Qy 3052 aacctagaacacatagttgactttataaaaagcagcgtaccacaccttggtgtgacag 3111
Db 3756 AGCTTTGAAATATTGTGAGTTTCTTAAAAACACGCTACCTGATATGAATACCTCTGAA 3815
Qy 3112 atgaaagacacatcaatcaggtatttgaaatggacatcgctaaacagtttacaagcttat 3171
Db 3816 ATGCAAAAATATTACCAGGTTTGTGAGATGATATTTCTAAGCAGTTGCTGATGCTAT 3875
Qy 3172 gaagtgtgaccagctctctcgaagaagaacttatcgattctctctctcctcagt---gac 3228
Db 3876 GAGGTGAATATCATGTGCTACAGGATGAGCTTCAGGAATCTTCATATCTCTGTGAGGAT 3935
Qy 3229 aaccaaagaatggaataattagaaacacacacagcagcttaacgaacacacacacacac 3288
Db 3936 AGTGAAGCTTTGGAGAGCTGGAGGCGCCCAATAGCAACTGGAAGACAAACATGGAC 3995
Qy 3289 ctcttgaacagttgcaggtggcaaatggttaggatccaaagccttgaggccaccattgag 3348
Db 3996 CTCCTAGAAAATTTACAGGTAGCTCATCTAAATCCAGGCTTTGGAATCAAACTCGAA 4055
Qy 3349 aagctctcagcagtgagagcaagctgaagcagggccatgcttaccttagcagtgagcgg 3408
Db 4056 AATCTTTTGGAGAGACACCAATGAAGTCTTTTAATCGGAGCCCTCGGAACACAGAAAAA 4115
Qy 3409 tcgcccctcgtcagcagcgtggagagcgtcgccgcgcg 3447
Db 4116 ATGCTTTATCAAAAGACAGTGGACCAACTCGGAAGCTG 4154

RESULT 10
AC009595/c standard; DNA; HTG; 189811 BP.
XX AC009595;
AC AC009595;
SV AC009595.3
XX
XX
XX
XX
DT 30-AUG-1999 (Rel. 60, Created)
DT 03-JUL-2000 (Rel. 64, Last updated, Version 4)
XX
XX
DE Homo sapiens chromosome 4 clone RP11-392K14 map 4, WORKING DRAFT SEQUENCE,
DE 28 unordered pieces.
XX
XX HTG; HTGS_DRAFT; HTGS_PHASE1.
XX
XX Homo sapiens (human)
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
XX [1]
RN 1-189811
RA Birren B., Linton L., Nusbaum C., Lander E., Allen N., Anderson M.,
RT "Homo sapiens chromosome 4, clone RP11-392K14",
RL Unpublished.
```

```
XX
RN 1-189811
RA Birren B., Linton L., Nusbaum C., Lander E., Allen N., Anderson M.,
RA Baker J., Baldwin J., Barna N., Beckerly R., Benn J., Brown A., Castle A.,
RA Carney J., Cavonius M., Collins S., Collamore A., Cooke P., DeArellano K.,
RA Depayre E., Devon K., Dewar K., Donelan L., Doyle M., Ferreira P.,
RA Fitzhugh W., Forrest C., Funke R., Gage D., Galagan J., Gardyne S.,
RA Gilbert D., Grant G., Hagos B., Heaford A., Horton L., Howland J.C.,
RA Jones C., Kann L., Karatas A., Lehoczy J., Lieu C., Locke K.,
RA MacDonald P., Marquis N., McEwan P., McGurk A., McKernan K., McLaughlin J.,
RA Melidrom J., Molla M., Morris W., Morrow J., Mychaleckyj J., Naylor J.,
RA Niloff M., O'Connor T., O'Donnell P., Pavlin B., Peterson K., Pollara V.,
RA Riley R., Roberts D., Roy A., Severy P., Stange-Thomann N., Stojanovic N.,
RA Stone C., Subramanian A., Tesfaye S., Tortuella-Miller I., Vassiliev H.,
RA Vo A., Wagner A., Wheeler J., Wu X., Wyman D., Ye W.J., Zody M.;
RT
RL Submitted (27-AUG-1999) to the EMBL/GenBank/DBJ databases.
RL Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,
RL Cambridge, MA 02141, USA
XX
CC On Jun 23, 2000 this sequence version replaced gi:6006228.
CC All repeats were identified using RepeatMasker:
CC Snit, A.F.A. & Green, P. (1996-1997)
CC http://ftp.genome.washington.edu/RM/RepeatMasker.html
CC ----- Genome Center
CC Center: Whitehead Institute/ MIT Center for Genome Research
CC Center code: WIBR
CC Web site: http://www-seq.wi.mit.edu
CC Contact: sequence.submissions@genome.wi.mit.edu
CC ----- Project Information
CC Center project name: L1755
CC Center clone name: 392_K14
CC ----- Summary Statistics
CC Sequencing vector: M13; M7815; 100% of reads
CC Chemistry: Dye-primers-amersham; 5% of reads
CC Chemistry: Dye-terminator Big Dye; 95% of reads
CC Assembly program: Phrap; version 0.960731
CC Consensus quality: 164704 bases at least Q40
CC Consensus quality: 175594 bases at least Q30
CC Consensus quality: 181024 bases at least Q20
CC Insert size: 214000; agarose-fp
CC Insert size: 187111; sum-of-contents
CC Quality coverage: 3.0 in Q20 bases; agarose-fp
CC Quality coverage: 3.5 in Q20 bas.
CC * NOTE: This is a 'working draft' sequence. It currently
CC * consists of 28 contigs. The true order of the pieces
CC * is not known and their order in this sequence record is
CC * arbitrary. Gaps between the contigs are represented as
CC * runs of N, but the exact sizes of the gaps are unknown.
CC * This record will be updated with the finished sequence
CC * as soon as it is available and the accession number will
CC * be preserved.
CC
CC 1 451: contig of 451 bp in length
CC * 452 551: gap of 100 bp
CC * 552 2663: contig of 2112 bp in length
CC * 2664 2763: gap of 100 bp
CC * 2764 4124: contig of 1361 bp in length
CC * 4125 4224: gap of 100 bp
CC * 4225 6767: contig of 2543 bp in length
CC * 6768 6867: gap of 100 bp
CC * 6868 8750: contig of 1883 bp in length
CC * 8751 8850: gap of 100 bp
CC * 8851 11272: contig of 2422 bp in length
CC * 11273 11372: gap of 100 bp
CC * 11373 13411: contig of 2039 bp in length
CC * 13412 13511: gap of 100 bp
CC * 13512 16962: contig of 3451 bp in length
CC * 16963 17062: gap of 100 bp
CC * 17063 22544: contig of 5482 bp in length
CC * 22545 22644: gap of 100 bp
CC * 22645 26284: contig of 3640 bp in length
CC * 26285 26384: gap of 100 bp
CC
```



```
CC * 26385 31223: contig of 4839 bp in length
CC * 31224 31233: gap of 100 bp
CC * 31324 36014: contig of 4691 bp in length
CC * 36015 36114: gap of 100 bp
CC * 36115 41939: contig of 5825 bp in length
CC * 41940 42039: gap of 100 bp
CC * 42040 48122: contig of 6083 bp in length
CC * 48123 48222: gap of 100 bp
CC * 48223 52983: contig of 4761 bp in length
CC * 52984 53083: gap of 100 bp
CC * 53084 58094: contig of 5011 bp in length
CC * 58095 58194: gap of 100 bp
CC * 58195 64606: contig of 6412 bp in length
CC * 64607 64706: gap of 100 bp
CC * 64707 71079: contig of 6373 bp in length
CC * 71080 71179: gap of 100 bp
CC * 71180 79886: contig of 8707 bp in length
CC * 79887 79986: gap of 100 bp
CC * 79987 86155: contig of 6169 bp in length
CC * 86156 86255: gap of 100 bp
CC * 86256 95370: contig of 9115 bp in length
CC * 95371 95470: gap of 100 bp
CC * 95471 103712: contig of 8242 bp in length
CC * 103713 103812: gap of 100 bp
CC * 103813 113565: contig of 9753 bp in length
CC * 113566 113665: gap of 100 bp
CC * 113666 123867: contig of 10202 bp in length
CC * 123868 123967: gap of 100 bp
CC * 123968 132296: contig of 8329 bp in length
CC * 132297 132396: gap of 100 bp
CC * 132397 141110: contig of 8714 bp in length
CC * 141111 141210: gap of 100 bp
CC * 141211 161622: contig of 20412 bp in length
CC * 161623 161723: gap of 100 bp
CC * 161723 189811: contig of 28089 bp in length.
```

XX Key Location/Qualifiers

```
FT source 1. .189811 /Chromosome="4"
FT /db_xref="taxon:9606"
FT /organism="Homo sapiens"
FT /map="4"
FT /clone="RP11-392K14"
FT /clone_lib="RP11 Human Male BAC"
FT misc_feature 1. .451
FT 552. .2663 /note="assembly_fragment clone_end:T7 vector_side:right"
FT misc_feature 2764. .4124 /note="assembly_fragment"
FT misc_feature 4225. .6767 /note="assembly_fragment"
FT misc_feature 6868. .8750 /note="assembly_fragment"
FT misc_feature 8851. .11272 /note="assembly_fragment"
FT misc_feature 11373. .13411 /note="assembly_fragment"
FT misc_feature 13512. .16962 /note="assembly_fragment"
FT misc_feature 17063. .22544 /note="assembly_fragment"
FT misc_feature 22645. .26284 /note="assembly_fragment"
FT misc_feature 26385. .31223 /note="assembly_fragment"
FT misc_feature 31324. .36014 /note="assembly_fragment"
FT misc_feature 36115. .41939 /note="assembly_fragment"
FT misc_feature 42040. .48122 /note="assembly_fragment"
FT misc_feature 48123. .52983 /note="assembly_fragment"
```

```
FT misc_feature 48223. .52983 /note="assembly_fragment clone_end:SP6 vector_side:left"
FT misc_feature 53084. .58094 /note="assembly_fragment"
FT misc_feature 58195. .64606 /note="assembly_fragment"
FT misc_feature 64707. .71079 /note="assembly_fragment"
FT misc_feature 71180. .79886 /note="assembly_fragment"
FT misc_feature 79987. .86155 /note="assembly_fragment"
FT misc_feature 86256. .95370 /note="assembly_fragment"
FT misc_feature 95471. .103712 /note="assembly_fragment"
FT misc_feature 103813. .113565 /note="assembly_fragment"
FT misc_feature 113666. .123867 /note="assembly_fragment"
FT misc_feature 123968. .132296 /note="assembly_fragment"
FT misc_feature 132397. .141110 /note="assembly_fragment"
FT misc_feature 141211. .161622 /note="assembly_fragment"
FT misc_feature 161723. .189811 /note="assembly_fragment"
FT XX
SQ Sequence 189811 BP; 52455 A; 40839 C; 40321 G; 53485 T; 2711 other;
```

Query Match 13.1%; Score 465.8; DB 30; Length 189811;
Best Local Similarity 99.6%; Pred. No. 5.2e-91;
Matches 467; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 416 aagtcctgagatcatcagctccatccgtcagcgagggaagatcgccggcagagagc 475
DB 97313 AGGTGCTGAGATCATCAGTCCATCCGACGGGGGAGATCGCCCGCAGCAGGAGC 97254
QY 476 tgcactgcccgtccgaggttcgacgacacgttttccaaagattcgaggtctttcg 535
DB 97253 TGCACGTGCGTCCGAGTTCGACGACACGTTTCCAAAGATTGAGAGTCTTCTGCG 97194
QY 536 gccgctgagctggcgacaaagaagctccgcccctgacgagtgatcgatcgaga 595
DB 97193 GCGCGTGCCTGGCGCCACAAAGAGCTCCGCGCCCTGTATCGACGAGTCGATCGAGA 97134
QY 596 agtcaatcacgtcagcgagccggggtccgagagagcccccagcccccagccatg 655
DB 97133 AGTTCATCAGTCAGCGGCGGCGGTTCCGAGAGAGCCCGCCCGCCCATG 97074
QY 656 ccgcccacagggagccaggagcctgtgagagcccatgagagccattcttccagc 715
DB 97073 CCGCGCCACAGGAGCCAGGAGCCTGTGCGAGGCCATGCGCAAGTCTTCTCCAGC 97014
QY 716 ccgcccagctgctgctgagccttaggaagagctgagagctgggggctccgagcag 775
DB 97013 CCGGCTGCGCTGCTGGCTTTAGGAAGAGCTGGGAGGCTGGGCGCTCCGAAGCAGCG 96954
QY 776 gcttctcagctccttcgagagagcagacattgagacacacctcattagcgacacaata 835
DB 96953 GCTTCTTCAGCTCCTTCGAGGAGAGCGACATTTGAGAACCCACCTCATTTAGCGACACATA 96894
QY 836 ttgtcagcccccagatcatcgaggaaataatcgaaactatgctcttcacat 884
DB 96893 TTGTGACGCCACAGATATCGAGGAAATCGAACTATGCTCTTTCACGGT 96845
```

RESULT 11
AC021106
LOCUS AC021106 195108 bp DNA linear PRI 03-JUL-2001
DEFINITION Homo sapiens clone Rp11-177C12, complete sequence.

AC021106 AC021106.6 GI:14589687
VERSION HTG.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 195108)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 195108)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 3 (bases 1 to 195108)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jul 3, 2001 this sequence version replaced gi:14018126.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/Index.shtml
Center project name: H_NH0177C12
----- Location/Qualifiers -----
FEATURES
source
1. 195108
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-177C12"
BASE COUNT 53372 a 40112 c 41873 g 59751 t
ORIGIN
Query Match 13.1%; Score 465.8; DB 9; Length 195108;
Best Local Similarity 99.6%; Pred. No. 5.2e-91;
Matches 467; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 416 aagtcgtcagatcagtcacatccatccgtcagcggggaagatgcgccgagcagagc 475
Db 180759 AGTGTCTGTAGATCATCATGCTCCATCCGTCAGCGGGGGAAGATCGCCCGCAGGAGGAGC 180818
Qy 476 tgcactgccctcaggtcagcagacacagcttttccaagaagttcaggtgctcttcgcg 535
Db 180819 TGCACGTGCCCTCGAGTTCGACGACAGCTTTTCCAAAGATTGAGGTGCTCTCTGCG 180878
Qy 536 gccgcgtgacgtggtgcacagaaggtctccgcggccctgacgagtgatcagcaga 595
Db 180879 GCCGCGTACGCTGCGGCACAGAAGGCTCCGCCGCCCTGATCGACGAGTGCATCGAGA 180938
Qy 596 agttcaatcagtcagcagcgccgcccggggtccgagagcccccgcacaccccccacatg 655
Db 180939 AGTTCAATCAGTCAAGCGGCGAGCGGGGGTCCGAGAGCCCCGCCCAACCCGCCCATG 180998
Qy 656 ccgcgcacacagggagcagagcctgtgcagcgcccatcgcaagtccttctccacgc 715
Db 180999 CCGGCGCCACAGGAGCCAGAGGCTGTGCCAGGCCCATCGCAAGTCTCTCTCCACGC 181058
Qy 716 ccggcctcgcctcgtcgtgcttttaggaaggagctgcaggtgggggctccgaagcagcg 775
Db 181059 CCGGCTCGCTCGCTGCGCTTTAGGAAGGAGCTGCAGGATGGGGCTCCGGAAGCAGCG 181118
Qy 776 gcttcttcagctcttcagagagcagcatttgaaacccactcattagcgagacacata 835
Db 181119 GCTTCTTCAGCTCTTCGAGGAGCGGACATTTGAGAACCACTCATTTAGCGGACACATA 181178

Oy 836 ttgtgcagccacagatcagagaaatcgaaactatgctcttccagat 884
Db 181179 TTGTGCAGCCACAGATATCGAGAAATCGAACTATGCTCTTCCAGGT 181227

RESULT 12
AC108933/c
LOCUS AC108933 208318 bp DNA linear HTG 01-FEB-2002
DEFINITION Homo sapiens chromosome 4 clone RP11-392K14, WORKING DRAFT
SEQUENCE, 5 unordered pieces.
ACCESSION AC108933 AC009595
VERSION AC108933.1 GI:18464316
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 208318)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 208318)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (01-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Feb 1, 2002 this sequence version replaced gi:8671945.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/Index.shtml
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: H_NH0392K14
Drafting center: WIBR
----- Summary Statistics -----
Sequencing vector: M13; %
Chemistry: Dye-terminator Big Dye; % of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 205652 bases at least Q40
Consensus quality: 206608 bases at least Q30
Consensus quality: 207022 bases at least Q20
Insert size: 218000; agarose-fp
Insert size: 207918; sum-of-contigs
Quality coverage: 6.45 in Q20 bases; agarose-fp
Quality coverage: 6.62 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 25746: contig of 25746 bp in length
* 25846: gap of unknown length
* 55131: contig of 39285 bp in length
* 55231: gap of unknown length
* 65232 105028: contig of 39797 bp in length
* 105029 105128: gap of unknown length
* 105129 159307: contig of 54179 bp in length
* 159308 159407: gap of unknown length
* 159408 208318: contig of 48911 bp in length.
FEATURES
Location/Qualifiers
1. 208318
/organism="Homo sapiens"
/db_xref="taxon:9606"


```
/db_xref="GI:1373163"
/translation="MRPAKRGKRDAAELRELWRTAIRQITMLNRMETENAMLOARON
ENELKRIKLYEEIIVPCDKQIIEERQIIEERNSTQIGNKKDKPKVLGHAIKRTGVPKSR
GDVMTFLAEQHSMTAPVDTKRFNFNTPYHMLKHLHTEHQHAFIDLGRTPFNHQFY
KDPGLGLOSLFNLKAYSLIDPELGYCGVLLGHLLHCDSEANSFOLLKHLMFRR
NMRTKYLPMKDKVLQVLSRLVKDHLPLDLYVWLDONDVSPITYAAAPWILTVFSOF
PLGVARVDFLLFTLESSDVIKFAIALLSVHKQOLLKDNDFEEMDYLKTVPKMEHT
CMEQIKLVFSMDIGKQLAEYNEYNVLOBELITNNHLEMLNREKTONQHLQOLQF
AQSIALETTRSSQAOITTLQKQSVQSLLEITQTLGRYVQLVEHPDLLEPNVEVR
MLQDLDDRRORPIPIETERIGKSVSNHSLGFLPKVLEELTERDELGSFKOKKEK
LPFEQLGQQQQHRLANGGQSSNVGSGSPPTPPSRNRLIDNASARTVMQVOKDELK
LPEHVDKRVANIKSLPEVDGVTPLPSPSTASNSGSIIFSRMGYRTTPPALSLAQ
ROSKYVATTPACQPHMEEVAPATTMAVPOEDVEEPQWPLPSMVGSDVNRKGT
OLKSIRPVHHRRAPIPLGGVQHPSTSEPAVRVAPVPELAPPAATATGGRS"
BASE COUNT      862 a 798 c 774 g 780 t
ORIGIN

Query Match      8 6%; Score 304.6; DB 3; Length 3214;
Best Local Similarity 56.0%; Pred. No. 6.1e-56;
Matches 621; Conservative 0; Mismatches 479; Indels 9; Gaps 2;

Qy 2155 gagctccgagagctggaagaagcttcttcacagatactgctgttagaagag 2214
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 169 GAGCTGAGGGAACCTGTGGCGACTGCCATTAGCGAGACTAATGCTGAACCGCATGGAG 228

Qy 2215 aaggaaaaatcagaagctccaaagcctctctgaaatgattgtgaaacaagcgcgtgaagctc 2274
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 229 ACGGAGAACGCCATTGTGCGAGCAGCGGACGAGATGAGACGAGCTAAAGCGCATTAACCTG 288

Qy 2275 gattatgaagaataactcctctgtcttaagaagaactacagttggaagaagatgctt 2334
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 289 GACTACGAGGAGATTGACCCCTGCGACAAAGCAGCTAATCGAGCGATGGGAGCAGATCAT 348

Qy 2335 ---agcactccgggaagatcaaaaaataagtttgatggaagaaatgcactcgctgtt 2391
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 349 GAGCGCAACTCCACACAGATAGCAACAGAGGATCCCAAGTCTGGCGACGCCATT 408

Qy 2392 gggcaagtggtgccacgtcaccgaggtgaaatctctggaatattctgctgagcaatc 2451
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 409 CGCACCGGAGTACCGCGATCGAAGCGCGCGATGCTGGACCTTCCTGGCTGAGCAGCAT 468

Qy 2452 cactctaa-----acaccagtttcccgaaacagcagcgaagatgtgcacacaaa 2505
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 469 TCCATGACACGGCACCGGTGGACAAAGCGATTCCCACTTCAATACACCGTATCAC 528

Qy 2506 gaactcttaagcagctgactctccagcagcagcagcagcagcagcagcagcagcagc 2565
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 529 ATGCTGCTGAACACCTTACTGAGCATCAGCATCGGATTTTCAATGATGTGGCGAGGACT 588

Qy 2566 ttctctacacacccatactctctccagcttgagcagcagcagcagcagcagcagcagc 2625
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 589 TTTCTTAATCACCACTTACAAAGATCCGCTTGTCTCGCGCATTTGCTGCTGTTTAAAC 648

Qy 2626 atttgaagggcctactcactcttagaccagaagtggaatattgccaaggtctcagcttt 2685
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 649 CTGCTGAGGGCTACTTCATTTAGATCCGGAATTTAGGGTACTGCGAGGGTCTGGGCTTC 708

Qy 2686 gtacgaggcatgttctctcttctcatatgattgaggaggcgttttaaaatgctcaagttt 2745
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 709 ATCTGCGGGCTTACTCTCTGCTGATTCGATGAAGCAATTCATTTCAACTGCTGAAGCAC 768

Qy 2746 ctgatgtttgacatgggctcgggaacacagtatcggcagacatgattattttacagatc 2805
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 769 CTAATGTTTTTCGCGGTAATATATGCGACGAATACCTTGGCGGACATGAAAAAGTTTCAACTG 828

Qy 2806 cagatgtaccagctctcaggttgccttatattacacagacagcagcagcagcagcagc 2865
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 829 CAGCTGTATCAGCTCTCCGATTTGTCAGAGGATCATCTGCCAGATCTCTACGTGTGGCTC 888

Qy 2866 gaggagcagcagatcggcccccagcctctacgtctgcccctgttctcaccatgtttgcc 2925
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 889 GATCAGACGATGCTCTCGCCCACTTTGTATGCGGGCTCCATGATGATCCTCACCGCTCTTTAGC 948
```

```
Qy 2926 tcacagttcccgctgggatttgtagccagagctctttgatgatgattttcttcagggaaca 2985
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 949 TCACAGTTTCGGCTGGGATTCGTGGCCCGCGCTCTTTTGATCTGCTCTTCTTCTGGAATCCTCC 1008

Qy 2986 gaggtcattataagtgagctttaagctctgttgggaagccataagcccttgattctgcag 3045
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1009 GATGTAATCTTTAACTTTTTCGCAATTCCTTCTCTCCGTGCACAAAGCAGCACTACTGCGC 1068

Qy 3046 catgaaaacctagaaccattagttgactttataaaaagcagcagcagcagcagcagcagc 3105
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1069 AAGGATAACTTCGAAGACATATATGACTATCTTGAAGACCGTGTGTCACCAAGATGGAGCAC 1128

Qy 3106 gtacagatggaaaagacacatcaatcaggtatttgaaaaggacatcgtctaaaacagttaca 3165
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1129 ACGTGCATGGAGCAAAATTAATGAAGCTGCTCTTACATGGAGCATTTGGAAGCAGCGTTGCC 1188

Qy 3166 gcttatgaagttgagtcacagctcttcaagaagaacttatcgattctctctctcagtt 3225
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1189 GAATACATATGGAGTACAAATGCTTTCAGGAGGAGATTACCACCACCTAACCATCACCTA 1248

Qy 3226 gacacccaagaatggataaattagagaa 3254
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1249 GAAATGCTTAACCGGGAGAGACGCGAGAA 1277
```

```
RESULT 15
LOCUS DME17919 4746 bp mRNA linear INV 18-NOV-1998
DEFINITION Drosophila melanogaster mRNA for pollux protein.
ACCESSION Y17919
VERSION Y17919.1 GI:3893102
KEYWORDS pollux gene.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 4746)
AUTHORS Xu,X.Z., Wes,P.D., Chen,H., Li,H.S., Yu,M., Morgan,S., Liu,Y. and
            Montell,C.
            Retinal targets for calmodulin include proteins implicated in
            synaptic transmission
            J. Biol. Chem. 273 (47), 31297-31307 (1998)
REFERENCE 99030403
MEDLINE 2 (bases 1 to 4746)
AUTHORS Montell,C.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-1998) C. Montell, The Johns Hopkins University
            School of Medicine, Department of Biological Chemistry, 725 N.
            Wolfe Street, Baltimore, MD 21205-2185, USA
```

```
FEATURES
source
    1..4746
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
    1..3316
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /map="83C"
        /tissue_type="retinal"
        /dev_stage="adult"
        452..4591
            /gene="pollux"
            /gene="pollux"
            /note="calmodulin binding protein"
            /codon_start=1
            /product="Pollux protein"
            /protein_id="CAA76939.1"
            /db_xref="GI:3893103"
```

```
gene
CDS
    1..4591
        /translation="MAELMHQMRDPAHTLGGVSGSIPQTLIGGGGGSHGNSGNALNGI
            HAYPATNLKMSAMNAOHDTSANPVSKMKAKSKSYTHGLSSSGTWNITPTSAQSN
            LSLADISPNIHTEFFVNYVKIKVSKRPNTFIDDALPKFAYDAQRLRLQLNRKM
```


